

This content has been downloaded from IOPscience. Please scroll down to see the full text.

Download details:

IP Address: 159.226.171.251

This content was downloaded on 09/09/2019 at 14:42

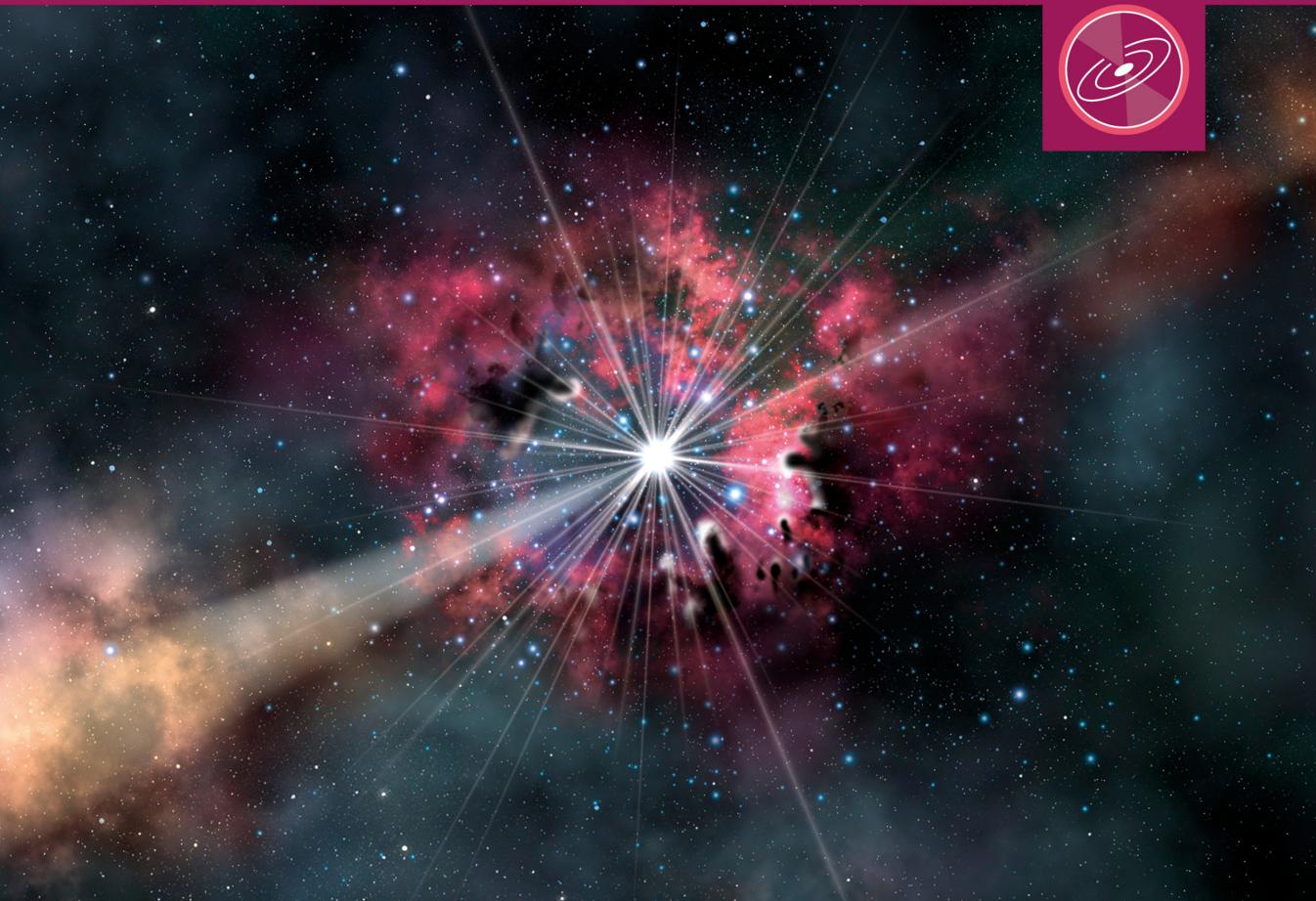
Please note that terms and conditions apply.

Gamma-Ray Bursts

Andrew Levan



American Astronomical Society



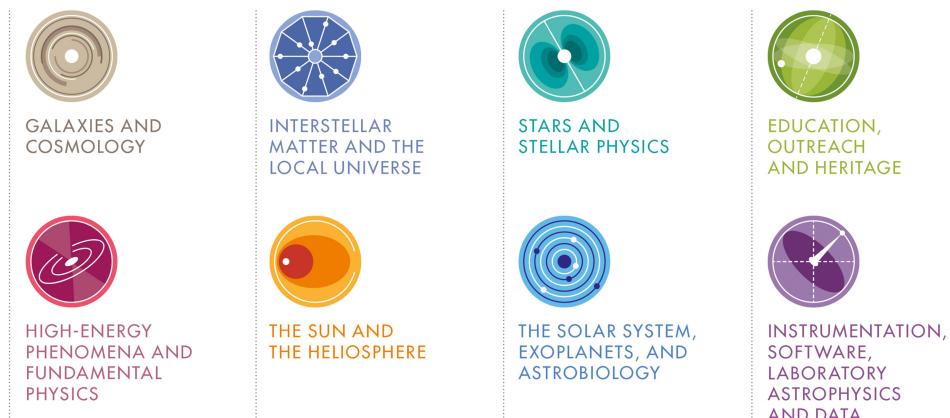
Gamma-Ray Bursts

AAS Editor in Chief

Ethan Vishniac, John Hopkins University, Maryland, US

About the program:

AAS-IOP Astronomy ebooks is the official book program of the American Astronomical Society (AAS), and aims to share in depth the most fascinating areas of astronomy, astrophysics, solar physics and planetary science. The program includes publications in the following topics:



Books in the program range in level from short introductory texts on fast-moving areas, graduate and upper-level undergraduate textbooks, research monographs and practical handbooks.

For a complete list of published and forthcoming titles, please visit iopscience.org/books/aas.

About the American Astronomical Society

The American Astronomical Society (aas.org), established 1899, is the major organization of professional astronomers in North America. The membership (~7,000) also includes physicists, mathematicians, geologists, engineers and others whose research interests lie within the broad spectrum of subjects now comprising the contemporary astronomical sciences. The mission of the Society is to enhance and share humanity's scientific understanding of the universe.

Gamma-Ray Bursts

Andrew Levan

Department of Physics, University of Warwick, UK

IOP Publishing, Bristol, UK

© IOP Publishing Ltd 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, or as expressly permitted by law or under terms agreed with the appropriate rights organization. Multiple copying is permitted in accordance with the terms of licences issued by the Copyright Licensing Agency, the Copyright Clearance Centre and other reproduction rights organizations.

Permission to make use of IOP Publishing content other than as set out above may be sought at permissions@iop.org.

Andrew Levan has asserted his right to be identified as the author of this work in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

Multimedia content is available at <http://iopscience.iop.org/book/978-0-7503-1502-9>.

ISBN 978-0-7503-1502-9 (ebook)

ISBN 978-0-7503-1500-5 (print)

ISBN 978-0-7503-1501-2 (mobi)

DOI 10.1088/2514-3433/aae164

Version: 20181201

AAS–IOP Astronomy

ISSN 2514-3433 (online)

ISSN 2515-141X (print)

British Library Cataloguing-in-Publication Data: A catalogue record for this book is available from the British Library.

Published by IOP Publishing, wholly owned by The Institute of Physics, London

IOP Publishing, Temple Circus, Temple Way, Bristol, BS1 6HG, UK

US Office: IOP Publishing, Inc., 190 North Independence Mall West, Suite 601, Philadelphia, PA 19106, USA

For Neil Gehrels. For showing how science could and should be done, and whose work in the field through BATSE, Fermi, and, in particular, his own Swift mission provides more content for this book than anyone else.

Contents

Preface	xii
Author biography	xiv
1 A Historical Primer	1-1
1.1 A Lesson in Serendipity	1-1
1.2 GRB Phenomenology	1-2
1.3 The Early Years	1-4
1.4 Suggested Models for GRB Creation	1-5
1.4.1 The Great Debate: Galactic or Cosmological	1-5
1.5 Intensive Efforts and Large Samples	1-6
1.5.1 The Angular Distribution of GRBs	1-6
1.5.2 Fluence or Luminosity Distributions of GRBs	1-8
1.5.3 Duration Distribution of GRBs	1-8
1.6 The Fireball Shock Model	1-10
1.7 The Long-GRB Afterglow Revolution	1-10
1.8 Redshifts and Host Galaxies	1-12
1.9 The Supernova Connection	1-13
1.10 GRB Energetics	1-15
1.11 The Neil Gehrels <i>Swift</i> Era	1-17
1.11.1 Short GRB Afterglows	1-18
1.11.2 Short GRB Progenitors	1-19
1.11.3 High-redshift GRBs	1-20
1.11.4 New Classes of GRB	1-23
1.12 New Insights from <i>Fermi</i>	1-25
1.13 Multimessenger Astronomy	1-26
1.13.1 Gravitational Waves	1-26
1.13.2 High-energy Neutrinos	1-29
1.14 Summary	1-29
References	1-30
2 Prompt Emission	2-1
2.1 Observational Properties	2-1
2.1.1 Temporal Structure	2-1
2.1.2 Spectral Structure	2-4
2.1.3 Polarization	2-7

2.2	Origin of the Prompt Emission	2-8
2.3	Summary	2-9
	References	2-10
3	Afterglow Emission	3-1
3.1	The First Afterglow Searches	3-1
3.2	X-ray Afterglows	3-6
	3.2.1 The Canonical X-ray Afterglow	3-6
3.3	Optical Afterglows	3-7
	3.3.1 Dark GRBs	3-8
	3.3.2 Polarization	3-9
3.4	Radio/Submillimeter Afterglows	3-10
3.5	Emission Processes	3-10
	3.5.1 Deceleration and External Shocks	3-10
	3.5.2 Synchrotron Emission	3-11
	3.5.3 The Reverse Shock	3-12
3.6	Evidence for Relativistic Beaming	3-14
	3.6.1 Jet Structure	3-15
	3.6.2 Simulations	3-17
	References	3-18
4	Central Engines	4-1
4.1	The Requirement of a Central Engine	4-1
4.2	Black Hole Central Engines	4-2
4.3	Magnetar Central Engines	4-4
4.4	Central Engines in Other Astrophysical Transients	4-7
4.5	Summary	4-10
	References	4-10
5	Long-GRB Progenitors	5-1
5.1	The GRB–supernova Connection	5-1
	5.1.1 The Collapsar Model	5-4
	5.1.2 Long GRBs without Supernovae	5-6
5.2	Observational Constraints on Stellar Masses and Sizes	5-7
5.3	Other Populations of Long-duration GRBs	5-9
5.4	Low-luminosity GRBs	5-9
5.5	Extremely Long Gamma-Ray Transients	5-10

5.6	Constraints for GRB Production	5-12
5.7	Binary or Single?	5-13
	References	5-17
6	Short-GRB Progenitors	6-1
6.1	Introduction	6-1
6.2	Progenitor Models	6-2
6.2.1	Creation of DCO Binaries	6-3
6.2.2	Mass Loss and Natal Kicks	6-6
6.2.3	Gravitational-wave Radiation-driven Mergers	6-8
6.2.4	Properties at Formation and Delay-time Distribution	6-9
6.3	Prompt Emission Properties	6-10
6.4	Afterglow Properties	6-12
6.5	Host Galaxy Properties	6-13
6.6	Locations	6-16
6.7	Redshifts and Energetics	6-18
6.8	Radioactively Driven Transients	6-19
6.8.1	Production of Kilonovae	6-19
6.8.2	Observational Constraints on Kilonovae	6-19
6.8.3	Heavy-element Production	6-21
6.9	Gravitational-wave Emission	6-22
	References	6-23
7	GRBs as Cosmological Probes	7-1
7.1	A Range of Cosmological Probes	7-1
7.2	Science from High- z GRB Afterglows	7-4
7.2.1	Selection of High- z GRBs	7-4
7.2.2	Metallicities and Hydrogen Column Densities	7-6
7.2.3	Line-of-sight Extinction	7-7
7.2.4	Constraints on Reionization	7-8
7.3	GRBs beyond $z \sim 5$	7-10
7.4	GRBs from Population III Stars	7-10
7.5	The Universal Star Formation Rate	7-11
7.6	Cosmological Parameters from GRBs	7-12
7.6.1	Prompt Emission Relations	7-13
7.6.2	Afterglow–Prompt Relations	7-16
7.7	The GRB Hubble Diagram	7-17
	References	7-17

8 Long-GRB Host Galaxies	8-1
8.1 Early Observations	8-2
8.2 GRB Hosts in the Galaxy Zoo	8-3
8.3 Basic Properties of Long-GRB Hosts	8-6
8.4 Building Meaningful Samples of GRB Hosts	8-6
8.5 GRBs Hosts at Optical and IR Wavelengths	8-8
8.5.1 Photometric Properties and Evolution	8-8
8.5.2 Host Galaxy Spectroscopy	8-10
8.5.3 Morphology	8-10
8.6 GRB Hosts at Submillimeter and Radio Wavelengths	8-12
8.7 GRB Hosts as Tools to Probe Progenitors	8-13
8.7.1 Metallicity Bias	8-13
8.8 GRB Hosts as Tools to Probe Distant Galaxies	8-14
8.8.1 Ly α Emission	8-14
8.8.2 The Host Galaxies of Dark and Dusty GRBs	8-15
8.8.3 GRBs and the Star Formation Rate	8-15
8.9 Burst Locations and Environments	8-16
8.9.1 Quantifying Burst Locations	8-16
8.9.2 Spatially Resolved Spectroscopy	8-18
8.10 Comparative Properties of GRB Hosts with Other Core-collapse Events	8-19
8.11 Summary	8-20
References	8-21
9 Multimessenger Astronomy	9-1
9.1 From Multiwavelength to Multimessenger Astronomy	9-1
9.2 Gravitational Waves	9-2
9.2.1 Gravitational-wave Detectors and Localization	9-3
9.2.2 Gravitational-wave Searches and Detection	9-6
9.3 Sources of Gravitational-wave Emission	9-8
9.3.1 Persistent Sources: Compact Object Binaries and Neutron Stars	9-9
9.3.2 Inspiral Sources: Compact Object Binaries	9-9
9.3.3 Bursting Sources: Core-collapse Events	9-10
9.4 Gravitational-wave Horizons	9-11
9.5 Prospect for Joint Detections	9-12
9.5.1 Gamma-Ray Bursts	9-13
9.5.2 Kilonovae	9-14
9.5.3 Orphan Afterglows	9-14

9.6	Electromagnetic Searches in Black Hole–Black Hole Mergers	9-15
9.7	GW 170817 and GRB 170817A	9-18
9.8	Gravitational Wave–Electromagnetic Detections: Questions for the Future	9-24
9.8.1	Future Prospects for Gravitational-wave Detectors	9-27
9.9	Neutrino Emission	9-28
9.10	Ultra-high-energy Cosmic Rays	9-31
9.11	Summary	9-32
	References	9-32
10	GRB Astronomy: Summary and Future Outlook	10-1
10.1	Challenges for the Future	10-1
10.1.1	The Identification of a Well-studied Sample of GRBs at High Redshift	10-1
10.1.2	The Detection of GRBs from First-generation Stars	10-2
10.1.3	The Use of GRBs to Determine Cosmological Parameters	10-3
10.1.4	The Detection of More GRBs in Coincidence with Gravitational-wave Triggers and the Addition of Neutrinos	10-3
10.1.5	The Direct Detection of GRB Progenitors	10-3
10.1.6	The Full Characterization of the GRB Blast Wave and Its Geometry	10-4
10.2	Possibilities for Future GRB Detection Missions	10-4
10.3	The Crucial Role of Follow-up	10-8
10.4	Summary	10-8
	References	10-9

Preface

The hook that is used to get somebody’s attention when talking about gamma-ray bursts (GRBs) is almost invariably that they are the most powerful explosions in the universe. Various impressive comparisons follow: they briefly outshine every other object in the visible sky, they create as much energy in a few seconds as the Sun does in its entire lifetime, the brightest events turn a significant fraction of the rest mass of the Sun into energy (just using $E = mc^2$), and many others. These are all true and no doubt contribute to the general public interest in GRBs, but there is much more astrophysical utility that arises from GRBs. They are not just interesting because they are so bright—the brightness of GRBs is a big part of what makes them extremely useful.

First, their luminosity means they can be seen across the universe, from the distant reaches, looking back to when the universe was only a few hundred million years old (5% of its current age), to galaxies where the light travel time was only a hundred million years or so in our cosmic backyard. When they explode, GRBs signpost these galaxies and illuminate their environments. The long-duration bursts are related to the collapse of massive stars. These stars live fast and die young, while their lower-mass cousins are still in their infancy. This means that long GRBs provide a way of tracing star formation. They are formed only from an exclusive and rare breed of star, meaning that their study enables unique constraints on stellar evolution. Short GRBs are created from the merger of two compact objects (neutron stars or black holes) and provide vital constraints on the evolution of binary stars, as well as being the source of the recently detected gravitational waves. The GRB itself (long or short) is created by the motion of material at very close to the speed of light and its interaction with both itself and the ambient medium.

These properties mean that GRBs provide direct information across a broad range of more “traditional” astronomical disciplines, including cosmology, galaxy evolution, stellar and binary evolution, the interstellar and intergalactic medium, relativistic astrophysics, multimessenger astronomy, and more. It is the questions that they answer and the connections that they provide across the breadth of contemporary astronomy that make them of interest to the public and many professional astronomers alike.

The aim of this book is to summarize our current understanding of GRBs by providing a historical overview of progress to date, as well as covering more details of individual aspects of GRB study, describing the consensus view of the phenomena, and highlighting the central questions that remain problematic. Since GRBs touch on so many different areas of astronomy, increasing numbers of researchers may need to learn something about them. The hope is that this book provides some of the background necessary for this and a stepping-off point for more detailed investigations. It is primarily aimed at PhD students and more established researchers, although I also hope that undergraduates with an interest in high-energy astrophysics may find it valuable (in particular Chapter 1).

This book is written by an observer from a predominantly observational perspective; because of this, the theoretical elements covered are relatively basic and, indeed, could cover a book on their own. The interested reader is particularly recommended to look at the in-depth review of Pawan Kumar and Bing Zhnag (2015, *PhR*, 561, 1) or the somewhat older work of Peter Meszaros (2002, *ARA&A*, 40, 137) for more detailed coverage of several of the more theoretical issues. Indeed, throughout, it is likely that my own bias as to the most interesting areas of study shines through, and I hope it does not grate too much with those who have other views. In fact, this is an e-book, which means it is more straightforward to make updates to it than in traditional print examples, so if you feel that there is a major omission, then please let me know, and perhaps it can be included in a subsequent version. Beyond this, GRB science has been a phenomenally fast-moving field, and it seems rather likely that updates will be required soon!

Finally, I'd like to add some words of thanks to those who have worked with me over the years, from my own introduction to the field of GRBs in 2001 to the present day. Particular thanks go to Nial Tanvir, Andy Fruchter, Melvyn Davies, Jens Hjorth, and Chryssa Kouveliotou. I would also like to thank my own more local group, who have provided much of the interaction and enthusiasm that continues to drive me, so thank you to all those who have worked with me at Warwick, Karl Svensson, Rachel Tunnicliffe, Greg Brown, Charlotte Angus, Darren White, Joe Lyman, Klaas Wiersema, Ben Gompertz, Elizabeth Stanway, and Danny Steeghs.

Andrew Levan
Stratford-upon-Avon, May 2018

Author biography

Andrew Levan



Andrew Levan is a professor of Astronomy and Astrophysics at the University of Warwick, UK. He has spent most of his career focusing on the origins of gamma-ray bursts and their use as cosmic probes, utilizing a wide range of ground- and space-based telescopes across the electromagnetic spectrum. Educated at the University of Leicester as both an undergraduate (2002) and postgraduate (2006), he has also held positions at the Space Telescope Science Institute and the University of Hertfordshire. He won a PPARC postdoctoral fellowship in 2005 and the Philip Leverhulme Prize in 2011. He is currently supported through a European Research Council Consolidator grant, awarded in 2016. Recently, he was a honoree in the inaugural 2018 Blavatnik Awards for Young Scientists in the UK.

Gamma-Ray Bursts

Andrew Levan

Chapter 1

A Historical Primer

1.1 A Lesson in Serendipity

Historians may remember that the partial nuclear test ban treaty of 1963 was a vital step in the de-escalation of the Cold War. It began controlling the proliferation of nuclear weapons and their tests. However, astronomers will likely relate much more strongly to its role in the origins of gamma-ray astronomy, in particular as a route to the discovery of the cosmic gamma-ray bursts (GRBs). Originally envisaged as a complete ban on nuclear testing, compromises during negotiation led to a partial ban, largely due to concerns about the means of verifying underground nuclear tests. This partial ban therefore banned atmospheric tests, ending the spectacular but highly toxic formation of mushroom clouds. It also banned testing underwater or in outer space. Questions were naturally raised as to how such a ban could be policed, especially for tests conducted beyond the Earth, where the typical seismic signatures could not be observed. Both the United States and the Soviet Union decided that monitoring for the gamma-ray flash produced at the moment of detonation was an effective route to the detection of illicit tests. To this end, the US *Vela* and Soviet *Kosmos* satellites were launched. Each was equipped with rather rudimentary gamma-ray detectors that could identify significant increases in high-energy photons above the background rate. Since each individual detector provided no location information, multiple satellites were launched into low Earth orbit, the use of which provided a continuous view of the Earth's surface (for ground-based tests), as well as the ability to use Earth-blocking and time-of-flight information to provide a crude localization of any flashes seen. In particular, the telltale sign of a nuclear test was expected to be a strong, fast (millisecond) outburst originating from the Earth or a known celestial body.

It remains unclear if these satellites ever detected illicit nuclear tests, although it does seem probable that the *Vela* Incident of 1979 was due to an atmospheric nuclear test. However, the *Vela* satellites are far more famous—at least in astronomical circles—for their discovery of GRBs (see Figure 1.1). They were initially detected as brief flashes in detectors on board the *Vela* spacecraft and first

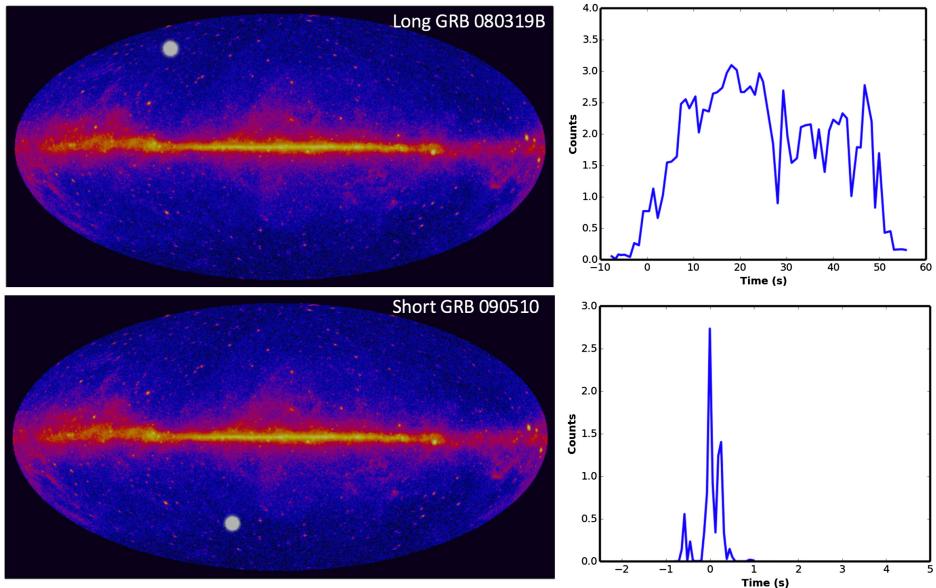


Figure 1.1. Simulations of the appearance of GRBs on the gamma-ray sky (background from the *Fermi* Large Area Telescope). The bursts appear as the brightest sources in the gamma-ray sky and then fade away. The right-hand panels show the light curves of the two GRBs. Animations are available online at <https://doi.org/10.1088/978-0-7503-1502-9>.

reported to the community in 1973 (Klebesadel et al. 1973; see Figure 1.2). Some events were seen simultaneously on the US *OSO-7* and *IMP-6* spacecraft (Wheaton et al. 1973), adding weight to their astrophysical reality, while a confirmation of their detection from the Soviet *Kosmos* spacecraft (*Kosmos-461*) was also rapidly forthcoming (Mazets et al. 1974). It was clear early on that these events did not meet the expectations of nuclear tests, while positional information (Wheaton et al. 1973) also showed that the events were not readily associated with local bodies such as the Earth, Moon, or other solar system objects. Such flares were completely unexpected in the community, although the discovery of cosmic X-ray emission from a similarly serendipitous route (via a sounding rocket aiming to detect solar X-rays scattered from the Moon) had already hinted at extremely energetic processes that might be occurring throughout the universe (Giacconi et al. 1962).

In the nearly half a century since the announcement of their discovery, GRBs have moved from a niche question to the astrophysical mainstream and are one of very few classes of sources to be identified from the very high energy gamma-ray to the low-frequency radio regimes.

1.2 GRB Phenomenology

A GRB is simply a burst of gamma rays, which can last from a fraction of a second up to several hours. Typically, GRBs appear from quiescence on timescales of a few seconds

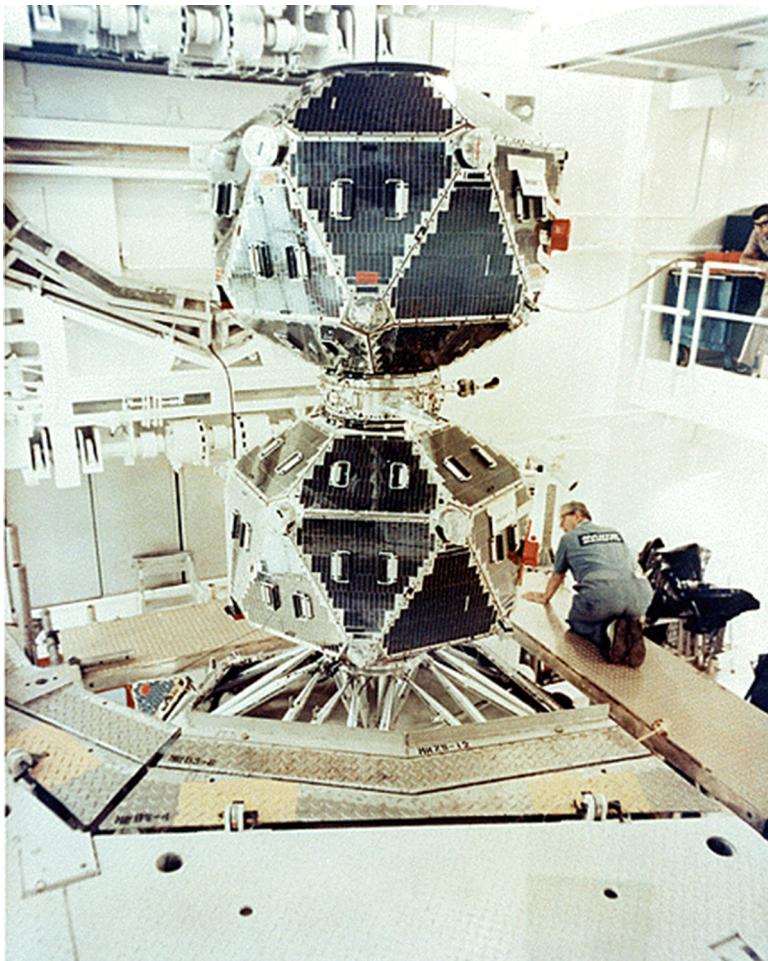


Figure 1.2. A US *Vela* satellite, responsible for the first detections of GRBs in the late 1960s. Multiple satellites were launched with the same design, enabling approximate positions to be derived from time delays between the detection of bursts in different satellites and Earth blocking. Image obtained from https://heasarc.gsfc.nasa.gov/Images/vela5b/vela5b_2.gif. Courtesy of NASA.

at most and then fade away again a few seconds later (a cartoon animation of long- and short-lasting GRBs is shown in Figure 1.1). In some cases, this is a single event, often described as a fast rise exponential decay (FRED). In other scenarios, the bursts show multiple rises and decays throughout the course of the prompt emission. The variability is sometimes rapid and has been seen on timescales as short as milliseconds. This, in turn, places stringent limits on the size of the region creating the GRB, since signals crossing the region cannot do so in less than the light travel time. For $\Delta T = 1 \text{ ms}$, $c\Delta T = 300 \text{ km}$, so the inferred region is only a few hundred km across, greatly restricting the range of astronomical bodies that could be responsible for GRB creation.

In general, there is significant diversity burst to burst, and few bursts look similar to others. An important feature is that GRBs do not repeat; they are one-off events,

after which the system creating them must either be destroyed or enter a quiescent period much longer than the lifetime of GRB-detecting missions (most likely the former is the case).

During the bursting phase, the apparent brightness of the bursts exceeds (sometimes by orders of magnitude) the brightness of any other source in the gamma-ray sky, and the brightest events are instantaneously brighter than the sum of all other gamma-ray emitters. It is particularly notable that GRBs emit the majority of their energy release in the gamma-ray regime. In particular, they emit most of their power at photon energies of tens to hundreds of keV,¹ while some have been detected at GeV energies (Abdo et al. 2009b). The actual flux output of a burst can be compared by looking at its energy spectrum, often referred to as the νF_ν or EF_E spectrum.² Doing so shows that GRBs have a clear spectral peak energy, often in the range of hundreds of keV. The vast majority of all the energy then produced is emitted in the X-ray and gamma-ray regime, and little is radiated in the more commonly observed optical window.

1.3 The Early Years

Through the late 1960s and early 1970s, GRBs were detected through scintillation detectors installed on satellites. Some of these were designed to search for gamma-ray emission from nuclear tests, while others were intended to track background radiation levels or measure high-energy emission from the Sun. They were not designed for the study of GRBs, and, combined with the limited sensitivity, this meant that only a handful of such bursts were detected. The first GRB catalogs from this period contained tens to perhaps 100 bursts (Klebesadel et al. 1973; Strong et al. 1974; Klebesadel et al. 1982) and in most cases are limited to light curves in a range of different energy bands, making direct comparison extremely challenging.

However, even these early data enabled significant insight into the GRB phenomenon. The light travel time arguments meant that theories of GRB creation needed to concentrate on compact regions, either individual dense stars (e.g., neutron stars or black holes) or subregions of larger objects, such as the cores of massive stars or stellar coronae. Many models were forthcoming in the few years following the detections of the first bursts. Indeed, by the end of the 1970s, over 30 models had been proposed, only one of which represented a “prediction” of gamma-ray emission (Nemiroff 1994) in being published prior to the first detection of GRBs. These models largely came from three different families: those that were powered by accretion power, accreting mass onto a white dwarf, neutron star, or black hole (with various different origins for the accreted material, including comets and flares

¹ In high-energy astrophysics, it is common to measure the relative wavelength or frequency of light as the associated photon energy, hence $E_\gamma = h \cdot c / \lambda$, and so a 1 keV photon has a wavelength of $\lambda \sim 1 \text{ nm}$.

² Since fluxes have units of, e.g., $F_\nu - \text{erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1}$, these cannot be directly compared, as they are also a function of the frequency being measured; hence, an optical F_ν might be much higher than one observed in gamma rays but does not actually emit as much energy, since it is emitting its energy at a far smaller frequency. Multiplying by the frequency enables the energies to be directly compared. However, note that to formally get the correct energy, you should multiply over the frequency range of observations (i.e., $\Delta\nu F_\nu$), not just the direct frequency, and this introduces offsets, typically of a few in such analyses.

from companion stars; Harwit & Salpeter 1973; Lamb et al. 1973); those related to stellar activity (e.g., directed stellar flares; Brecher & Morrison 1974); and those due to the catastrophic destruction of stellar-sized objects. However, this is by no means an exhaustive list of the possibilities that have been discussed, which are as extreme as GRBs originating from white holes or cosmic strings (neither of which have actually been identified). Indeed, a comprehensive list of models up until 1992 published by Robert Nemiroff (1994) contains a total of 118 different models for the creation of GRBs. Remarkably, many of the different physical mechanisms suggested in this list have subsequently been shown to occur in the universe and may well result in observable electromagnetic emission. However, the vast majority do not create GRBs. Indeed, it is striking that the model commonly invoked today to explain the origin of most bursts is not on this list, as it was not published until 1993 (Woosley 1993).

Since the error regions determined by triangulation methods relying on satellites in Earth orbits were typically large, it was impossible to pinpoint the bursts on the sky. Early triangulation methods (Klebesadel et al. 1973; Wheaton et al. 1973; Strong et al. 1974) did show that there was no obvious preference for bodies sufficiently rare that they could be isolated from a handful of events; for example, it was clear that GRBs were not originating exclusively from any of the planets or the galactic center. However, the typical error regions contained many thousands of galactic stars and background galaxies, any of which could have hosted the GRB. Hence, these locations were not sufficient to provide a means of distinguishing between the ever-increasing number of models for GRB production.

One solution to this problem was to increase the baseline used in triangulation. By doing this, the differences in the time of flight for photons became increasingly longer, and so the region to which the burst can be located became commensurately smaller. There was a major motivation in the creation of the InterPlanetary Network (IPN).³ Since gamma-ray detectors are relatively simple and not too heavy, it was possible to place them on numerous spacecraft. Importantly, this included not only those slated for orbit around Earth but also those destined for the farther reaches of the solar system. With baselines >1 au, it would, in principle, be possible to get positions that were orders of magnitude better than previously possible. The hope was that these precise positions would be the decisive step toward to origin of GRBs.

1.4 Suggested Models for GRB Creation

1.4.1 The Great Debate: Galactic or Cosmological

Through the late 1980s, the absence of precise GRB positions precluded the identification of their nature, and the controversy over the origin of GRBs intensified. Two camps made strong arguments between Galactic and extragalactic models. While many models fell by the wayside in this period, those involving neutron stars either within or outside the Milky Way continued to gain traction.

³ <https://heasarc.gsfc.nasa.gov/docs/heasarc/missions/ipn.html>

In 1995, a “great” debate on the issue was held, honoring an earlier debate in 1920 discussing the scale of the universe. While perhaps somewhat less all-encompassing, the scale of GRBs was clearly one of the major issues in astronomy at the time. In this debate, Bodhan Paczynski argued for a cosmological origin (Paczynski 1995), while Don Lamb presented the case for a galactic origin (Lamb 1995). The debate was heavily skewed by the recent announcements from the Burst and Transient Source Experiment (BATSE) about the isotropy of GRBs on the sky (see below and Briggs et al. 1996). Such isotropy naturally favors cosmological models, as it is inevitable in essentially all of them. However, galactic models remained possible thanks to the realization that neutron stars received so-called natal kicks on their formation (see also Chapter 6) and that these kicks could remove many neutron stars into a corona around the Galaxy. Both parties agreed that further observations were necessary, noting the importance of measuring redshifts, either from the bursts themselves or from counterparts identified at other wavelengths (Rees 1995). No doubt this debate and its ensuing publicity were important in energizing these efforts throughout the late 1990s.

1.5 Intensive Efforts and Large Samples

Throughout the 1980s, more efforts were made to understand the nature of GRBs, and their origin became one of the headline questions in astrophysics. Recognizing this newfound importance, their study rapidly became a motivation for new astrophysical satellites, perhaps most notably as part of NASA’s Great Observatories program. Probably the most impactful series of scientific instruments of all time, these satellites—the *Compton Gamma-Ray Observatory* (*CGRO*), the *Hubble Space Telescope* (*HST*), the *Chandra X-ray Observatory* (*CXO*), and the *Spitzer Space Telescope*—have revolutionized our view of the universe across the electromagnetic spectrum. While each of them have made major contributions of the field of GRBs, it is the *CGRO* that made the most immediate impact following its launch. This is because of the presence of BATSE on board the satellite. BATSE consisted of 16 scintillation gamma-ray detectors, with two placed on each corner of the instrument. One of these was optimized for burst detection, providing a large field of view and a route to determining directionality. The other was designed to enable higher-energy resolution and better spectroscopy of the detected bursts.

BATSE delivered crucial diagnostics into bursts thanks to two vital improvements. First, the sensitivity of the detectors meant that for the first time, bursts were able to be detected in large enough numbers and with a sufficiently well-understood selection function that statistical studies gained much more power. Second, the error regions obtain for BATSE bursts, while very large for typical optical telescopes (several degrees in diameter), were sufficient to identify any signatures of large-scale structure on the sky.

1.5.1 The Angular Distribution of GRBs

Perhaps the most important result from BATSE was that GRBs appeared to be located isotropically on the sky—they came from no preferred direction (Figure 1.3).

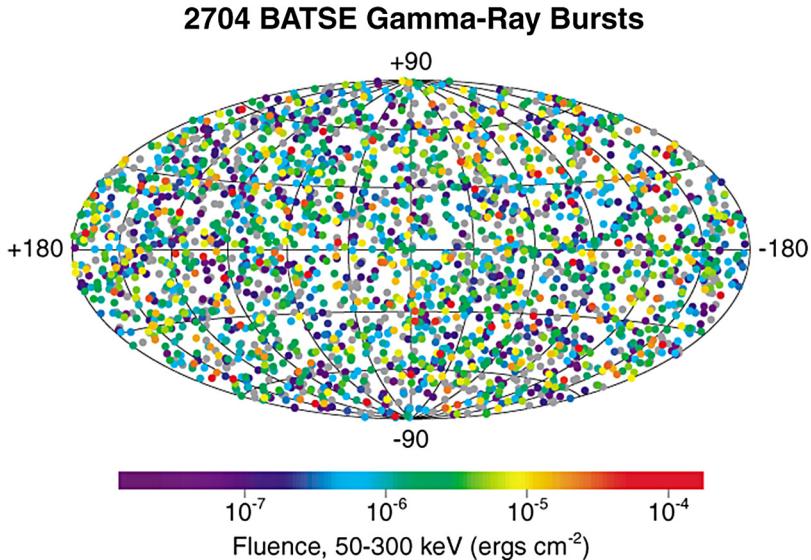


Figure 1.3. Distribution of BATSE GRBs on the sky, color coded by the measured fluence of the burst (Briggs et al. 1996). The bursts are distributed isotropically on the sky and show no preference for any direction. This is true at any fluence level. Image obtained from <https://heasarc.gsfc.nasa.gov/docs/cgro/batse/>. Courtesy of NASA.

This is extremely important, as there is a very limited range of distances over which isotropic signatures would be expected. For example, bursts arising in the solar system may be expected to fall preferentially on the ecliptic, bursts within the Galaxy should be concentrated toward the galactic plane, and bursts in very local galaxies (within a few Mpc) may well be identified as concentrations of bursts toward those particular directions. The isotropy of GRB positions immediately ruled these locations out. Indeed, it left essentially two possibilities. Bursts could be located at distances of perhaps a few hundred pc, within the thin disk of the Milky Way. If their visibility horizon was of this scale, then we would not be able to view the decline in density in directions out of the plane of the Galaxy or see far enough into the disk to observe strong correlations with the galactic plane. The second possibility was that GRBs were in fact cosmological. Since the cosmological principle states that on large scales, the universe is isotropic and homogeneous, the distribution of bursts was naturally explained in this scenario as well.⁴

⁴Formally, there may also be locations in the Galactic halo where such distributions could also have been observed; however, this would be problematic because it would require stellar-scale events that could only occur in halo systems and never in the stellar population of the disk or bulge. Since the bulge is also made of old stars, this rapidly becomes problematic. Since the Galaxy halo also extends to large radii, the absence of any overdensity in the direction of M31 would also be problematic for this interpretation.

1.5.2 Fluence or Luminosity Distributions of GRBs

The large number of bursts detected by BATSE also allowed the distribution of the total fluence of the bursts to be studied.⁵ Of particular value is the so-called $\log N - \log S$ distribution, which describes the number of bursts (N) of a given fluence (S). It is sometimes also considered as the $\log N - \log P$ relation, where the P refers to the peak flux rather than the total fluence. While such distributions had been examined for GRBs before (Mazets & Golenetskii 1981), BATSE provided the first opportunity to observe such a distribution with a good number of bursts in a given fluence bin. The principle of the $\log N - \log S$ test is simple. Imagine a set of standard candles distributed uniformly in space. As the distance to the source increases, the number of sources would increase with the volume included (essentially the cube of the distance). However, the relative fluence observed from each source would decrease following the expected inverse-square law for light ($1/\text{distance}^2$). Hence, the slope of the $\log N - \log S$ would be expected to be $-3/2$. While luminosity functions and cosmological evolution naturally impact this, the broad picture is the same. A $-3/2$ slope is indicative of a uniformly distributed sample, while something cut off at fainter values is more suggestive of a limited horizon over which the sources can be viewed.

The $\log N - \log S$ for GRBs shows the expected $-3/2$ slope for brighter bursts but deviates away at the fainter end, providing fewer faint bursts than expected (Paciesas et al. 1999). Taken at face value, this suggests that we do see to the edge of the GRB distribution. However, this does not provide information about where this “edge” may lie. It could be within the Galactic disk or halo, or perhaps it could be our universal horizon should the bursts be extremely energetic.

1.5.3 Duration Distribution of GRBs

The duration distribution of GRBs offers another handle on their properties and is perhaps one of the characteristics of the prompt emission that has proved most valuable in identifying multiple classes of GRB. The duration is not necessarily straightforward to define, since the differing morphologies of individual GRBs mean that the bursts can look very different. For example, some bursts exhibit multiple episodes of flaring activity, and others show only one. Some are fainter but longer-lived. There have been various approaches to determining both the duration and the variability of GRB light curves, but the duration distinction that has been most widely used is T_{90} , the duration over which 90% of the total fluence of a given burst is recorded. In essence, T_{90} can be obtained by integrating the observed GRB light curve and then determining the times within the light curve when 5% and 95% of the total fluence has been observed. This provides a measurement of T_{90} . Avoiding the regions where the burst is rising out of the background or fading back into it

⁵Fluence is the time integral of the flux over the duration of the burst, so it is a measure of the total energy recorded in the detector during the burst. If a distance is known, it is therefore a route to measuring the total energy of the burst. Hence, while flux is described as energy per unit time per unit area, fluence is an energy per unit area.

becomes a relatively robust measurement of the duration of the burst, although other, even tighter restrictions such as T_{50} (the duration over which 50% of the fluence is observed) are also used.

The duration distribution of BATSE GRBs is shown in Figure 1.4. In this figure, it is clear that there are two classes of GRB with median durations of about 0.5 and 30 s (Kouveliotou et al. 1993). When examining the spectral properties of these bursts, the distinction becomes even cleaner, and the shorter GRBs emit more high-energy emission than the longer bursts. This leads to the identification of two populations of GRBs: short hard bursts and long soft bursts. Indeed, evidence for these populations was apparent as early as the beginning of the 1980s (Mazets & Golenetskii 1981) but did not become widely acknowledged until the results from BATSE. Within this distribution there have also been claims, based on BATSE data and observations made more recently with *Swift*, of populations of intermediate-duration bursts ($2 \text{ s} < T_{90} < 10 \text{ s}$; Mukherjee et al. 1998), as well as ultra-long GRBs (ULGRBs) ($> 10^{3-4} \text{ s}$) (Levan et al. 2014). The reality of these as separate classes and their relation to either the long or short GRBs remains a matter of discussion.

It should be stressed that T_{90} , while powerful, is a blunt tool for measuring GRB durations. More sensitive detectors may track emission for longer, while bursts also have differing durations in different energy bands. Hence, the measurement is detector-dependent. It has recently been recognized that the dividing lines between different populations should therefore actually be different, depending on the detector in use (Bromberg et al. 2013), and that the distinction between long and short GRBs at 2 s is only approximate. Furthermore, T_{90} is always defined in the observer frame, not in the rest frame of the burst. In principle, this means that bursts observed at 2 s at some T_{90} at $z = 0$ would last for $T_{90} \times (1 + z)$ at some higher redshift. In practice, the evolution of duration is apparently much weaker than this because of detector effects (Littlejohns et al. 2013), but it may have a potentially important effect.

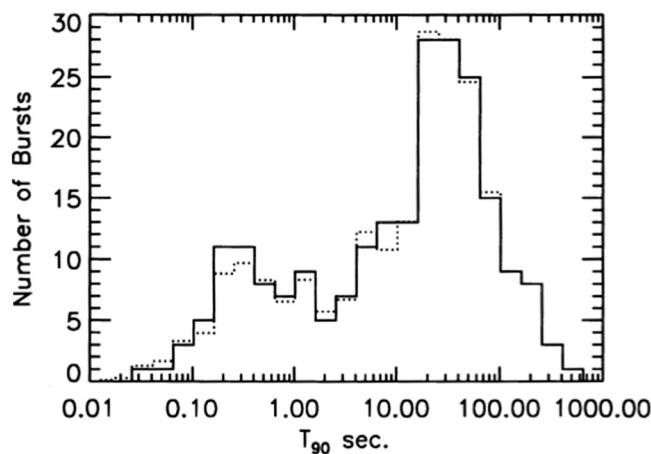


Figure 1.4. Duration distribution of BATSE GRBs from Kouveliotou et al. (1993). The two populations are clearly visible as short and long GRBs. © 1993. The American Astronomical Society.

1.6 The Fireball Shock Model

The evidence (in particular, the spatial distribution) gathered in the early 1990s lent credence to cosmological models of GRBs. In this scenario, the burst energies are particularly challenging, with isotropic energy releases (if at $z \sim 1$) of up to 10^{54} erg, comparable to a significant fraction of a solar rest mass. Combined with the inferred size of the emission region based on variability arguments, the energy density was extreme, placing this energy within a region only a few km across. This greatly limited the physical processes that were plausible for the creation of GRBs but also suggested that it may be possible to consider the observed properties in a progenitor-independent manner simply by considering what may happen to such energy density when released.

In addition to the extreme energy constraints, the gamma-ray emission itself also represented a puzzle—bursts apparently created many photons well above the pair-production threshold ($E = 0.511$ MeV), which should have created electron-positron pairs rather than the observed gamma rays, and the solution to the problem was not obvious.

It was solved with the development of the fireball shock model (e.g., Rees & Meszaros 1994). The basic premise of this model is that the energy deposition within a small volume drives a relativistic expansion from the source. This reduces the photon energy in the rest frame (where the photons are generated) by a factor of Γ^{-1} , where Γ is the bulk Lorentz factor of the outflow. Driving such a powerful outflow, reaching Lorentz factors of several hundred is possible if the initial explosion energy is very high and the outflow contains little matter—so-called weak baryon loading.

Since gamma-ray bursts exhibit significant variability, it is apparent that the source of the energy is, in general, not steady but instead variable. If the energy input at the base of the explosion varies, one might expect the Lorentz factor of material emitted at different times during the explosion to vary as well. In this case, as the material streams out at ultrarelativistic velocities, some ejecta will be moving more quickly than other parts of the ejecta. Eventually (probably after hundreds to thousands of seconds in the frame of the star but much less in the frame of the ejecta or the observed burst), ejecta emitted at different times will interact. These shocks can create gamma rays and could therefore explain the observed properties and energies of the GRBs.

This was an important breakthrough in the understanding of GRBs; but perhaps more importantly, the interaction of this outgoing relativistic flow with whatever material surrounded the star would potentially create much longer-lived emission, a so-called afterglow. Such an afterglow could be found at different wavelengths with much more sensitive instrumentation, offering the possibility of pinpointing GRBs on the sky and resolving questions of their distance and energetics. This prediction gave new impetus to attempts to identify GRBs outside the gamma-ray regime.

1.7 The Long-GRB Afterglow Revolution

The brevity of GRB emission was a significant hurdle in understanding the nature of GRBs. Not only was it nearly impossible to obtain highly accurate positions from

the gamma rays alone, even if such positions had been forthcoming, it would in general have been impossible to observe the locations with sufficient sensitivity while the burst was ongoing (indeed, even today, this has only been achieved on a handful of occasions for very bright and/or very long-lived bursts; Racusin et al. 2008; Greiner et al. 2014). For this reason, the possibility of identifying GRB afterglows was particularly appealing.

Motivated by this possibility, various groups initiated plans to search for afterglows. However, this was a particularly challenging prospect for BATSE error boxes of tens of square degrees. Indeed, attempts to locate afterglows to BATSE bursts were not successful (Krimm et al. 1994).

Instead, afterglows were located following the launch of *BeppoSAX*, a joint Italian and Dutch satellite. *BeppoSAX* (often more affectionately referred to as *SAX*) offered two crucial enhancements for the detection of GRBs. The first was a wide-field X-ray monitor that could deliver positions that were accurate to arcminutes, and the second was the ability to repoint and observe the source position within an on-board X-ray telescope on a timescale of a few hours. Both of these capabilities were crucial. The shift from degree- to arcminute-sized error boxes that were available soon after the GRB was a breakthrough for afterglow searches. Rather than requiring specialist instrumentation, these error regions could fit within the field of view of available CCD detectors on the majority of ground-based telescopes. Within a few arcminutes, there may also be only tens to hundreds of sources. Hence, searches for afterglows within these regions were finally tractable. Indeed, on timescales of days, X-ray positions that were even more accurate could be released.

Finally, with the burst of 1997 February 28, all of these ingredients were put into place. A long burst was detected by the gamma-ray detectors. The location was followed up by narrow-field instruments onboard *BeppoSAX* (Costa et al. 1997) and independently from the William Herschel Telescope on La Palma (van Paradijs et al. 1997). Both X-ray and optical observations were revisited several days later and revealed a fading source in both wavebands. This was the first detection of a GRB afterglow. This afterglow therefore pinpointed the source on the sky and provided a route to measuring the host galaxy (Sahu et al. 1997) and, ultimately, after 4 yr, the redshift for GRB 970228 (Bloom et al. 2001). In practice, redshifts were already obtained for several bursts by the time the redshift of GRB 970228 was found, with the first redshift measurement for a burst for GRB 970508, the same burst for which the first radio afterglow was discovered (Metzger et al. 1997; Frail et al. 1997).

This afterglow era has revolutionized the study of GRBs, and it is thanks to observations of afterglows and the science they enable that studies of GRBs have made such inroads into many fields of astrophysics. Despite the uncertainty over the behavior of afterglows at this time, the general model that was used for multi-wavelength follow-up has become the gold standard today. A burst is detected by a space-based mission, and it has its position rapidly telemetered to the ground while simultaneously pointing narrow-field instruments at the same location. These narrow-field instruments, in concert with ground-based observatories, then pinpoint the afterglow and measure the burst redshift.

The crucial difference between these early studies and observations today is the timescale on which the observations occur. In 1997, the process took several days. Now, with observations with *Swift*, it is routine for these observations to take seconds to minutes, with major advantages in efficiency and scientific return. Indeed, this formula is so successful that it is exactly the approach being proposed for the next generation of GRB detection missions, such as *SVOM* and *THESEUS* (see Chapter 10).

1.8 Redshifts and Host Galaxies

Once afterglows were discovered, it was possible to use them to precisely locate bursts on the sky and identify any underlying source. While this is now routine, it was not the case following the discovery of the first afterglow. Indeed, while deep observations did unveil an extended underlying source, it was less clear if this source was in fact a distant host galaxy or a more local nebula (Sahu et al. 1997). A claim of possible proper motion at first suggested that GRBs could indeed be galactic systems (Caraveo et al. 1997), since it would be impossible to observe proper motion from objects at cosmological distances.

The solution was, of course, to obtain a direct measurement, either from the afterglow or the underlying source. This was first done for the afterglow of GRB 970508, which exhibited weak absorption lines consistent with a redshift of $z = 0.835$. In other words, the afterglow was illuminating material at $z = 0.835$ somewhere along the line of sight between the burster and us. The most likely explanation for this is that we are observing material in the host galaxy of the GRB itself, and so the burst lies at $z = 0.835$. Formally, the presence of absorption lines only means that the burst lies at a redshift greater than that of the absorption. Indeed, in some cases, intervening absorbers are also seen. However, in most cases, the most distant absorber is ascribed to the host galaxy of the burst, and in the few cases where we can see lines varying, having been excited by the UV flash from the burst itself,⁶ this is indeed the case. This direct redshift measurement immediately rules out all galactic models for long-GRB production and implies that they are highly energetic cosmological explosions.

The explosions do not occur randomly in space but instead tend to land directly on host galaxies. In addition to measurements from the afterglow, the redshift can also be obtained from the host galaxies themselves; in most cases, this is possible through the presence of a series of emission lines, including the [O II] (3727 Å) doublet, [O III] (4959, 5007 Å), and hydrogen Balmer lines. These lines are primarily excited by star formation and are absent from elliptical or passive galaxies. Their presence in essentially all GRB hosts bright enough to be observed implies that the host galaxies themselves are star-forming. This is confirmed by extensive *HST* rest-frame UV imaging, which shows the host galaxies to be the bright UV sources, often with clumpy morphology suggestive of multiple star-forming regions within the host galaxies (Fruchter et al. 2006). Among long-GRB hosts, there are no examples of

⁶These are the so-called fine-structure lines that respond to the UV flash and then decay.

passive, non-star-forming host galaxies. The galaxies are typically of relatively low luminosity compared to either field galaxies or the hosts of other explosive events (such as core-collapse supernovae), and this in turn implies that while they do trace star-forming host galaxies, they are only tracing some fraction of the star formation. The preference for faint galaxies suggests that this may be due to a bias toward forming GRBs at lower metallicity due to the relation between stellar mass (or luminosity) and gas-phase metallicity (Tremonti et al. 2004). Such inferences were later confirmed by direct spectroscopic studies and deeper photometric work (Savaglio et al. 2009; Svensson et al. 2010; Hjorth et al. 2012; Perley et al. 2016).

1.9 The Supernova Connection

While observations of host galaxies strongly favor an origin for long GRBs in exclusively star-forming galaxies, this does not directly tie the bursts to the massive (supernova-producing) population of massive stars within them. For example, such young stellar populations also host significant numbers of young, and hence rapidly spinning, pulsars that might, in principle, produce bursts (Duncan & Thompson 1992; Hurley et al. 2005). Indeed, it is striking that flares from young neutron stars are now among the prime contenders for the origins of the fast radio bursts (FRBs; Lorimer et al. 2007; Thornton et al. 2013), which may have some connection to GRBs (Metzger et al. 2017). Hence, there was significant interest in ascertaining if the GRBs were directly associated with the core collapse of massive stars, and hence with supernovae.

The possibility that such connections could be directly tested was clearly demonstrated in May 1998, with the discovery of the burst GRB 980425. This long burst overlapped a local galaxy (ESO 184-G082) with a known redshift of $z = 0.0085$. Observations taken of the burst over the few days after its occurrence revealed not a fading afterglow but an apparently rising supernova, named SN 1998bw (Galama et al. 1998). This supernova was extremely bright compared to the majority of supernovae observed, peaking at an absolute magnitude of $M_B \sim -19.3$, a factor of 10 brighter than most core-collapse events and comparable to a Type Ia supernova. It was also spectrally unusual, showing no signs of either hydrogen or helium emission lines (a so-called Type Ic supernova) but indications of very high expansion velocities ($20,000\text{--}30,000 \text{ km s}^{-1}$) over the first few weeks. The presence of such a rare supernova, spatially and temporally coincident with a GRB, was extremely unlikely by chance and strongly suggestive of an association. However, such an association was nontrivial to interpret. Where was the afterglow? Was this very nearby burst, with an energy 10,000 times less than most GRBs, in any way related to the “normal” GRBs? Indeed, these questions remain relevant today, and GRB 980425 remains the closest GRB ever seen by some margin.

Importantly, the supernova in GRB 980425 provided a guide as to what could be observed in more distant and typical GRBs. The supernova was bright, perhaps in contrast to some models of so-called “collapsars” for the creation of GRBs, in which material would be accreted directly onto the newly formed black hole, and any supernova shock may be weak (or even absent). This meant that such bumps could, in theory, be detected by current technology out to $z \sim 1$. Such searches were

naturally undertaken, and it rapidly became apparent that many bursts had the decay of their optical afterglows halted and often reversed by a rising component, strikingly similar to the supernova seen in GRB 980425. Further observations moved beyond single colors and showed that this similarity extended into the spectral regime. The bumps not only had a similar shape and luminosity to SN 1998bw, they also exhibited the same colors. From this, a consensus gradually built that long GRBs were essentially always associated with broad-lined (\equiv high-velocity) Type Ic supernovae with a rather narrow range of peak brightness.

However, while this was true, a direct evolving spectral sequence for comparison to SN 1998bw remained elusive. Any such sequence could only be obtained for a local burst $z < 0.2\text{--}0.3$, and such bursts were naturally rare. The opportunity finally arose in 2003, with the detection of GRB 030329 at $z = 0.17$ (Figure 1.5). From this,

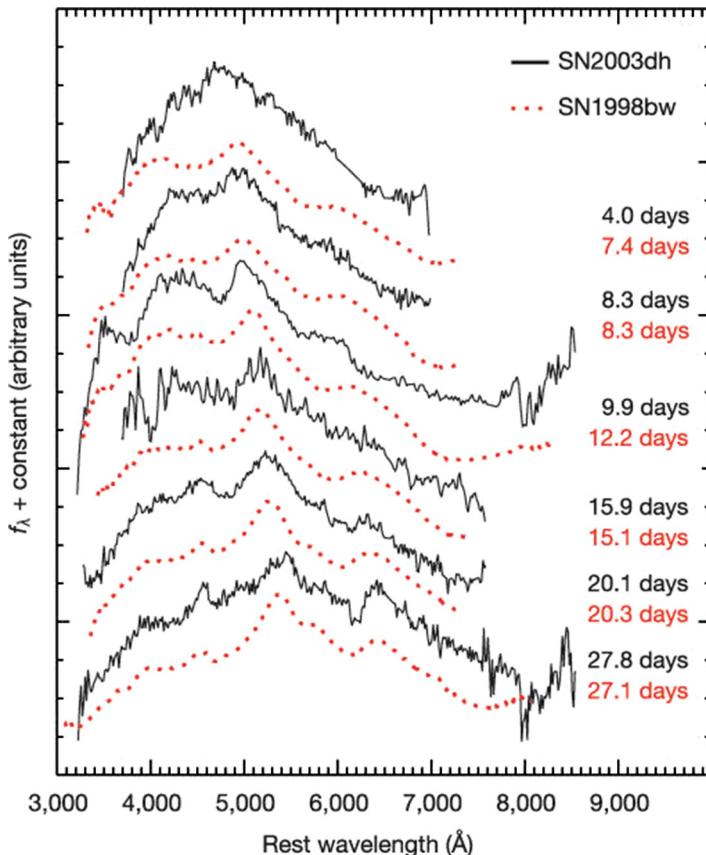


Figure 1.5. Spectral evolution of GRB 030329, associated with SN 2003dh (Hjorth et al. 2003). The solid lines show SN 2003dh, while the dashed red lines show SN 1998bw, associated with GRB 980425. The broad similarity in the two is strong evidence that highly energetic, cosmological GRBs are associated with stellar core-collapse events. Reprinted by permission from Springer Nature: Hjorth et al. (2003).

in-depth observations finally showed the emergence of spectral features that were nearly identical to those of GRB 980425 but in a GRB a factor of 1000 more energetic (Hjorth et al. 2003; Stanek et al. 2003). The supernovae associated with GRB 980425 (SN 1998bw) and GRB 030329 (SN 2003dh) remain exemplars for the GRB–supernova connection today.

Detailed observations of GRB supernovae remain challenging because of the rarity of local events and the difficulty in isolating the supernova light against contributions from the afterglow and underlying host galaxy. However, the above picture remains largely unchanged; the vast majority of long GRBs show supernovae of similar peak luminosity to SN 1998bw. This is true across the breadth of the GRB luminosity function. Indeed, perhaps the strongest connection between the faintest and brightest bursts arises from the very similar supernovae that appear in both examples. Only in a handful of cases is this picture complicated. Two notionally long GRBs apparently show no supernovae (Fynbo et al. 2006; Gal-Yam et al. 2006; Della Valle et al. 2006), although the nature of these events remains controversial (Gehrels et al. 2006). There is also one example of a ULGRB (Levan et al. 2014) in which the supernova appears perhaps a factor of >2 brighter than SN 1998bw (Greiner et al. 2015). While such events may ultimately prove highly instructive in elucidating the nature of the GRB central engine, it seems unlikely that they detract from the broader conclusions that GRBs arise from hydrogen- (and helium-) depleted progenitors.

1.10 GRB Energetics

At the distances now measured for GRBs, they are extremely energetic. A common metric for determining these energies is the isotropic equivalent energy release (E_{iso}); this is the energy that the burst would have if it emitted the energy observed by us to observers in all directions. It can be calculated from the measured fluence (S_{γ}) following

$$E_{\text{iso}} = S_{\gamma} \frac{4\pi d_L^2}{(1+z)}, \quad (1.1)$$

where d_L is the luminosity distance to the burst and z is the redshift. This differs slightly from the standard luminosity relation, which would omit the $(1+z)$ factor; since the fluence is the time integral of the flux and is integrated in the observer frame, dividing by $(1+z)$ correctly places it back in the emitting frame. A further subtlety arises for the use of S_{γ} , which is dependent on the energy range integrated over. In principle, one might like a bolometric (total) fluence, and bands such as 1–10,000 keV are often used. However, it is also common to use the energy range of the detector in question, for example, 15–150 keV in the case of the *Swift*-BAT, which can lead to underestimates of the total fluence (or alternative bolometric corrections) of a factor of 5–10.

At the typical redshifts of GRBs, the inferred isotropic energies of long bursts are in the region $10^{52} \text{ erg} < E_{\text{iso}} < 10^{54} \text{ erg}$. These are extreme energies; for an efficiency factor of $\epsilon \sim 0.1$, they are the equivalent of $m = E_{\text{iso}}/\epsilon c^2 \sim 1 M_{\odot}$. In other words,

they suggest that of order a solar mass of material is rapidly transferred into radiation on timescales of only a few seconds. This is extremely challenging for almost any model, and indeed, it does not seem to reflect the true energies of GRBs. The solution lies in the highly relativistic and beamed nature of GRBs. Relativistic expansion was already expected for GRBs as a natural component of the fireball model, so this creates relativistic beaming (where light is beamed in the direction of motion), although it was always also thought likely that GRBs would be geometrically beamed as well (i.e., the emission would be naturally confined to a jet with some narrow opening angle). Because of this beaming, they only illuminate a small fraction of the sky, which can be expressed through the beaming factor, $f_b \sim \theta_j^2/2$ (Frail et al. 2001), where θ_j is the half opening angle of the jet (the angle from the polar axis to the edge of the jet) in radians. The true energy of the GRB is then $E_{\text{true}} = f_b E_{\text{iso}}$.

Direct evidence for beaming in some GRBs can be found from achromatic breaks in the GRB light curve due to a combination of geometric and relativistic effects (see Chapter 3). The time of these breaks can be linked back to the opening angle of the jet, hence enabling the degree of collimation in a given GRB to be measured. For long GRBs, this suggests that collimation is typically limited to opening angles of a few degrees and beaming corrections of a factor of 100 or more. When correcting for this, it has been noted that the energies of GRBs are actually remarkably tightly correlated around 10^{51} erg, suggestive of rather similar energy injection into each event (Frail et al. 2001; Bloom et al. 2003). Short GRBs exhibit markedly different energetics. They are generally less energetic by ~ 2 orders of magnitude, and only a handful of examples demonstrate strong evidence for beaming (Fong et al. 2015).

It should be noted that there is also a potentially significant population of underluminous bursts that may have volume densities significantly higher. Because they are faint, they cannot be seen to such large distances, and this “Malmqvist” bias results in a far lower *observed* rate of low-luminosity events, despite a far higher *intrinsic* rate. These events have energies that more typically cluster in the range $10^{48} \text{ erg} < E_{\text{iso}} < 10^{50} \text{ erg}$, and they include some of the most famous and well-studied bursts, including many of those crucial for the supernova–GRB connection. The connection of these bursts to the more luminous events remains a subject of debate. It is now clear that both low- and high-luminosity GRBs can host very similar supernovae, indicating that they arise from the same physical systems, but the cause of the difference remains uncertain. It may well be that the low-luminosity systems are those where we observe the X/gamma-ray shock breakout from the star but may not directly see the highly relativistic jet (Nakar and Sari 2012).

An alternative route to obtaining GRB energetics arises from very late time observations, particularly in the radio band, where so-called calorimetry can be undertaken. At these late times, the blast wave can be treated as being essentially spherical, so concerns related to the beaming angle are largely removed, providing a route of converting directly from the observed flux to a total energy.

1.11 The Neil Gehrels *Swift* Era

Following the first detection of afterglows and the realization that GRBs could be located across the electromagnetic spectrum, thoughts naturally turned to routes in which the process of afterglow detection could be optimized. It was apparent that afterglows were fading sources, so it was expected that they should be brighter, and hence easier to detect, if observed at earlier times. The GRB missions in operation at the end of the 1990s provided positions based on the hard X-ray and gamma-ray detections, and observations to search for afterglows tended to then be conducted hours later, at least in the first months and years of the afterglow era. This led to the development of the *Swift* concept, a capable platform that would contain gamma-ray, X-ray, and UV/optical instrumentation and be able to autonomously slew rapidly to the locations of any detected GRB. *Swift* was selected for launch as a medium-sized explorer mission (Midex) in 1999 October, shortly after the observation by a ground-based robotic telescope of the extremely bright optical afterglow of GRB 990123 (Akerlof et al. 1999) had spectacularly demonstrated the benefits of early observations of GRBs. Developed rapidly over the next 5 yr, *Swift* successfully launched on 2004 November 20 and remains in operation. It has recently been renamed the *Neil Gehrels Swift Observatory* in recognition of its extremely influential Principal Investigator, Neil Gehrels (1952–2017).

Swift contains three main instruments: the Burst Alert Telescope (BAT), the X-ray Telescope (XRT), and the UV and Optical Telescope (UVOT); a full description of these instruments is given by Gehrels et al. (2004), but they are described briefly here. The gamma-ray detections come from the BAT, which uses a coded mask design, in which a mask with 54,000 randomly placed lead tiles is located above a set of detectors. These tiles cast shadows down onto the detectors, with the illumination pattern being a function of the location of a gamma-ray source relative to the instrument. Hence, the illumination pattern of the detectors, in concert with the known positions of the tiles, can be used to determine a position to any given gamma-ray source. The BAT has a field of view of 1.4 sr (approximately one-sixth of the sky) and an effective energy range of 15–150 keV, well suited to the spectral regions where GRBs emit much of their light. It typically yields gamma-ray positions for GRBs with error radii of 2'–4' (90% confidence). The BAT is complemented by the XRT, which is a narrow field-of-view ($23.6' \times 23.6'$) instrument with an effective energy range of 0.2–10 keV but a far higher sensitivity than the BAT, reaching $\sim 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ in 10^4 s, compared with a sensitivity of the BAT of $\sim 10^{-8} \text{ erg s}^{-1} \text{ cm}^{-2}$ (Gehrels et al. 2004). (Note that because GRBs, unlike their afterglows, only last a few seconds, integrating for longer with the BAT will not improve the sensitivity for a GRB.) This enables the detection of the X-ray afterglow, and X-ray positions for GRBs are now typically accurate to 1"–2" (Evans et al. 2009). At the same time as the XRT is observing, so is the UVOT. This is a 30 cm telescope working from ~ 1700 to 6000 \AA and is based strongly on the design of the *XMM-Newton* optical monitor. It contains a set of UV and blue optical filters and cycles through these on a given burst, reaching limiting magnitudes in each filter in the region of 20–22 in the few hours after the burst detection.

Swift has dramatically reduced the timescale for observations of GRBs. In the early days of GRB detections, observations on a timescale of 10 hr were considered to be a rapid response, but *Swift* points its narrow-field instruments at bursts on timescales of 60–100 s. Once a burst is detected by the BAT, a detection alert is sent directly to the ground through the military TDRSS system, triggering observations on the ground. At the same time, an autonomous decision is made onboard that decides if the burst has sufficient merit (based on a series of onboard figures of merit) to override the current observation and slew to the new GRB target. If this is done, the telescope slews and then points the narrow-field instruments at the burst location, typically 60–100 s after the first trigger, depending on the distance the telescope must slew. This provides rapid dissemination of X-ray and UV positions, and images based on short initial observations are also sent directly to the ground at this time to enable early enhanced positions to be obtained, as well as verifying the reliability of any afterglow identification (e.g., confirming it is not a cosmic ray, etc). This means that positions with $\sim 2''$ accuracy are now available for the majority of GRBs on timescales of 2–5 minutes after the initial detection.

The impact of *Swift* on the GRB field has been profound; not only has the number of bursts increased by an order of magnitude, from $\sim 1 \text{ month}^{-1}$ to $\sim 2\text{--}3 \text{ week}^{-1}$, but the vast majority of these bursts have X-ray afterglows and positions sufficiently accurate to locate either an optical afterglow or an underlying host galaxy. The afterglows have multiwavelength coverage beginning seconds after the burst trigger and continuing in some cases for several years. This combination of increased sensitivity, increased time range, and improved wavelength coverage has provided the route to a wealth of new discoveries that have been enabled by *Swift*. These are discussed in more detail below, as well as in relevant chapters throughout the text.

1.11.1 Short GRB Afterglows

Perhaps the major outstanding question in the GRB field at the launch of *Swift* was the nature of the short-duration GRBs. While afterglows had been discovered for several tens of long GRBs at the start of the *Swift* mission, none had been located for short bursts, despite intensive efforts targeted at the handful of reasonably localized short bursts. The pinpointing of short bursts and enabling the same diagnostics for short GRBs as had been possible for the long bursts was a priority.

The first opportunity arose on 2005 May 9 with the detection of the faint short-duration GRB 050509B. A prompt slew by *Swift* enabled it to catch a very faint and fading X-ray afterglow containing only a handful of X-ray photons (Gehrels et al. 2005). However, this was enough; while the error region enabled was still several arcsec in radius, it was sufficient to place the burst close to a massive elliptical galaxy in a merging galaxy cluster at $z = 0.225$ (Bloom et al. 2006). This was an immediate hint; a long GRB had never been seen in an elliptical galaxy, and this particular elliptical exhibited no signs of star formation to deep limits. Deep observations failed to uncover any evidence of supernova emission that should have been detected had the progenitor been a massive star (Hjorth et al. 2005). Furthermore, the location of the burst, while uncertain, was apparently well away from the core of the Galaxy.

Despite intensive efforts targeted at GRB 050509B, no optical afterglow was detected. Instead, it was the burst GRB 050709 (interestingly, a burst first detected by *HETE-2*) that yielded the first optical afterglow detection (Hjorth et al. 2005), and subsequently, rare radio afterglows have also been detected. It is apparent that these afterglows are systematically fainter than those of the long bursts, providing a natural explanation for their nondetection prior to the ability of *Swift* to provide rapid, accurate positions. This is not surprising, since the total fluence of short GRBs is also much lower than that of the long bursts, and since afterglow brightness broadly scales with fluence (Gehrels et al. 2008; Nysewander et al. 2009) the faintness is consistent with expectations.

Swift has now detected over 100 short GRBs, with approximately 70% having X-ray afterglow identifications. In the majority of these cases, observations have sought to identify optical afterglows, with approximately 25% exhibiting optical afterglows. These observations also frequently identify candidate host galaxies, and it is from these galaxies that the distances to almost all short GRBs are derived (Fong et al. 2013). These redshifts are predominantly at $z < 1$, with a mean redshift of $z \sim 0.5$ (D’Avanzo 2015). The short GRBs are well scattered on their host galaxies, often lying off the stellar light (Fong & Berger 2013), and in some cases potentially at such large offsets that host galaxy identification is nontrivial, so-called hostless bursts (Berger 2010; Tunnicliffe et al. 2014).

1.11.2 Short GRB Progenitors

The continued observations of the short GRB population back up the picture painted by studies of the first events and offer robust measures of, for example, the demographics of the host galaxies or the offset distribution. These observations are consistent with the long-favored model for the creation of short GRBs in the final gravitational wave–driven merger of either a neutron star–neutron star or neutron star–black hole binary. Since the merger time from formation of the second compact object scales with the separation at this time to the fourth power ($t_{\text{merge}} \propto a^4$), it is naturally expected that some binaries will have long merger times and thus occur in elliptical galaxies. Indeed, several of the known NS–NS binaries within the Milky Way have merger times longer than the age of the universe. Equally, somewhat smaller separations on formation can dramatically alter the merger time (e.g., a factor of 2 in separation yields a factor of 16 in merger time). so it is also expected that some binaries will arise from star-forming systems.

Beyond this, the neutron star binaries receive a kick in formation due to natal kicks to neutron stars that are known to be in the region of hundreds of km s^{-1} (Arzoumanian et al. 2002) and because of mass loss at the time of each supernova. Hence, not only do we expect to associate short GRBs from mergers with old populations, we expect to frequently find them at large offsets from their host galaxies. Hence, the galaxy properties and locations of the short GRB population are qualitatively in keeping with the expectations of this model. Indeed, when compared to population synthesis calculations that evolve many binaries and then move them within the potential of their host galaxies, a solid quantitative match is

found between the properties of the short GRBs and the expectations of compact object mergers (Church et al. 2011).

However, even stronger evidence for this origin arises from the identification of so-called kilonova⁷ emission in at least one short GRB, GRB 130603B (Berger et al. 2013; Tanvir et al. 2013). The principle of a kilonova, much like a supernova in long GRBs, is that radioactive material is launched from the merger site, and the decay of this material with time should power an optical transient that can be seen for longer than the GRB and probably longer than the afterglow (Li & Paczyński 1998). Hence, in much the same way that long GRBs have their afterglow decay slowed or even reversed by the presence of a supernova, short GRBs should exhibit mid-to-late time bumps in their light curves from kilonovae. The difference in the case of kilonovae is that rather than being powered by radioactivity in iron-group elements, in particular the decay of ^{56}Ni , kilonovae are powered by the decays of much heavier elements. These are the so-called *r*-process elements that are highly neutron-rich (hence why neutron star mergers are a promising location for their synthesis). Many of these elements also have very high optical opacities due to the complex electron shells around them. This means that while supernovae are predominantly optical events, kilonovae may release much of their light in the infrared (Barnes & Kasen 2013).

In the case of GRB 130603B (Berger et al. 2013; Tanvir et al. 2013), the optical afterglow apparently continued to fade following its early decay, while the infrared showed a much slower decay, with evidence for rebrightening at around 10 days (Figure 1.6) at a luminosity consistent with the expectations for a kilonova (Barnes & Kasen 2013). Subsequent reanalysis of other short GRBs showed some evidence for similar bumps (Jin et al. 2015; 2016), perhaps hinting that such emission could be universal. Unfortunately, kilonovae are a factor of >20 fainter than the supernovae associated with long GRBs and can only readily be discovered in local examples, with even *HST* reaching its limits before $z \sim 0.4$.

However, the detection of this fingerprint of neutron star mergers offered strong evidence for the origin of short GRBs and also for a possible signal to search for in the error boxes of gravitational-wave sources (see Section 1.9 and Chapter 9).

1.11.3 High-redshift GRBs

The exceptional luminosity of GRBs and their afterglows makes them powerful cosmological probes. In addition to the studies of the interstellar and intergalactic medium enabled by their afterglows (Fynbo et al. 2009), their brightness makes them ideal candidates for exploration of the universe at the highest of redshifts. Furthermore, since their progenitors are thought to be born from low-metallicity stars, their rate may well be higher (per unit star formation) at higher redshift (Bromm & Loeb 2006; Fruchter et al. 2006; Perley et al. 2016). Indeed, there have been suggestions that the very first generation of stars, pristine systems made only of

⁷ Kilonova emission is also variously referred to as mini-supernova or macronova emission.

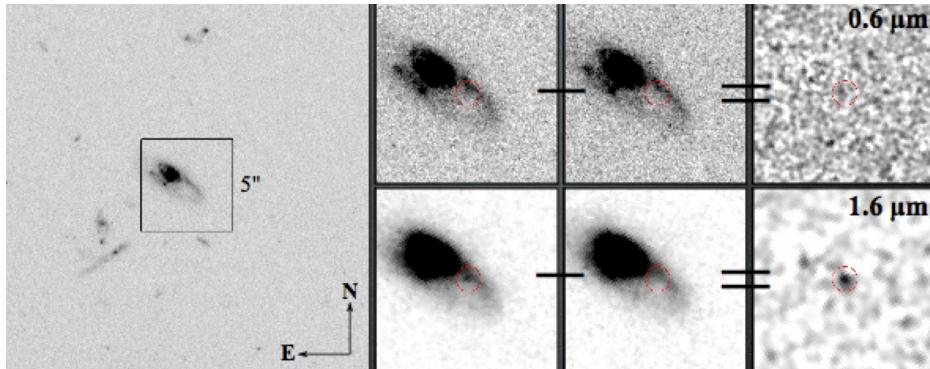


Figure 1.6. The *HST* observations of the counterpart of GRB130603B taken \sim 10 and \sim 30 days after the explosion. The left-hand image shows a wide-field view, while the top row shows optical and the bottom rows shows infrared observations of the counterpart. The final column shows the result of subtracting the last image from the first. As can be clearly seen, a residual is visible in the infrared, while none is seen in the optical. This is consistent with the presence of an infrared bump, which can be interpreted as a red kilonova (Tanvir et al. 2013; Berger et al. 2013). Reprinted by permission from Springer Nature: Tanvir et al. (2013).

hydrogen and helium (often referred to as Population III stars), could generate bright GRBs (Toma et al. 2011; Piro et al. 2014), perhaps making GRBs an ideal probe to locate the very first stellar collapse.

Observing GRBs at such high redshift is not trivial. The large luminosity distances involved mean that only the typically brighter bursts can be detected by current satellites, so the rate of bursts observed at such large distances does not simply follow the enclosed volume but is markedly truncated by instrumental sensitivity. Simulations suggest that only a small fraction, $<10\%$, of bursts observed by *Swift* are likely to lie at $z > 5$ (Bromm & Loeb 2006).

The discovery of high- z GRBs is potentially more straightforward than for other high-redshift sources. Their brightness means that they are readily identifiable on relatively short exposures on telescopes of modest apertures. In some cases, this enables the direct spectroscopy of their afterglows. The presence of metal absorption lines and frequently strong Ly α absorption then provides a route to redshift measurement. At higher redshifts, this becomes more challenging. This is not necessarily because the increasing distance makes the afterglows much fainter, since time dilation also means we observe them at earlier times, when they are intrinsically brighter (in practice, these two effects do not directly offset one another, and many high- z GRBs are relatively fainter). The main problem, which can also be an asset, is the presence of the Ly α break. The Ly α break arises because of the presence of neutral hydrogen in the line of sight between the GRB and ourselves. Photons emitted at wavelengths $<1216 \text{ \AA}$ (Ly α) can then interact with this neutral hydrogen, creating absorption features. Since there are many clouds, this creates a myriad of features known as the Ly α forest. Photons with wavelengths $<912 \text{ \AA}$ (the ionization energy of hydrogen, or so-called Lyman limit) are blocked altogether. At lower redshift, the Ly α forest is not too dense, and some light can be seen down to the

Lyman limit, creating a strong break in the spectrum at $912 \times (1 + z)$ Å. At higher redshifts, increasing amounts of light are absorbed in the forest shortward of Ly α and create a strong break at $1216 \times (1 + z)$ Å (this is true for $z > 5$). At even higher redshifts, where the universe becomes globally neutral (rather than just containing clouds of neutral gas), the absorption short-ward of Lyman- α becomes complete, creating a region of wavelength (equivalently redshift) space over which the absorption is total. This is known as the Gunn–Peterson trough. However, while breaks around Ly α can be readily identified based on photometry alone (see Figure 1.7), distinguishing between a strong Ly α forest and the Gunn–Peterson trough is only possible via high-quality spectroscopy.

Given the motion of the Ly α break with redshift, it can be seen that observations in the optical window, where many GRB afterglow searches are carried out, are simply not sensitive to high- z GRBs. For example, the UVOT on *Swift* can only observe light to ~ 5000 Å, so it cannot observe bursts beyond $z \sim 4.5$. However, while the Lyman break poses problems for some observations, it also provides a solution to ready redshift measurements. In particular, multiwavelength observations show a strong break in the spectrum at the position of Ly α in the rest frame. Identifying this break in photometry provides an immediate indication of high redshift and motivates more intensive efforts to obtain direct spectroscopic confirmation.

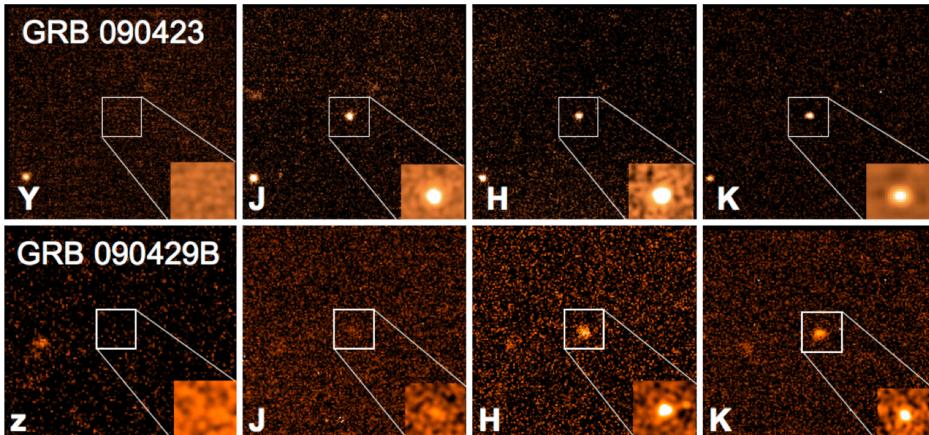


Figure 1.7. Selection of two high-redshift GRBs, GRB 090423 (Tanvir et al. 2009) and 090429B (Cucchiara et al. 2011). Wavelength runs to the right, with bluer images on the left. The afterglow is at the center of both figures. It can be seen to vanish in the bluest band while appearing in the redder bands. Importantly, it is also of comparable brightness in all the redder bands. This is indicative of an intrinsically blue source, with a strong break introduced because of the presence of neutral hydrogen clouds in the intergalactic medium, and identifies these two events as high-redshift bursts. The burst GRB 090423 has a subsequent spectroscopic measurement of $z = 8.23$ (Salvaterra et al. 2009; Tanvir et al. 2009), while GRB 090429B has its redshift fixed only by this photometry to $z \sim 9.4$ (Cucchiara et al. 2011). Reprinted by permission from Springer Nature: Tanvir et al. (2009).

This is exactly the route through which the most distant GRBs have been observed. For example, in 2005 September, the burst GRB 050904 was identified as an *i*-band dropout, indicating little light blue-ward of $\sim 7750 \text{ \AA}$ and a redshift of $z > 6.3$ (Haislip et al. 2006). Subsequent spectroscopy from Subaru was able to pinpoint the redshift to $z = 6.29$ (Kawai et al. 2006). A similar route was used in 2009 April for the burst GRB 090423, where observations from Gemini implied a dropout in the *Y* band at around $1 \mu\text{m}$ and a photometric redshift of $z > 8$. Spectroscopy from the TNG in La Palma and VLT in Chile was then able to confirm a redshift of $z = 8.2$ (Tanvir et al. 2009; Salvaterra et al. 2009; see also Figure 1.7). These bursts are of particular importance; for a time, GRB 090423 was the most distant spectroscopically confirmed object. While galaxies have now likely been found beyond this point, it still pinpoints that stars have had time to form and evolve in < 600 million yr from the Big Bang. Indeed, since the properties of GRB 090423 do not appear remarkable when compared to those of other GRBs, it may be reasonable to speculate that it did not form from the first, pristine generation of stars (so-called Pop III) and that 600 million yr is sufficient for at least two generations of star formation.

These GRBs also provide valuable constraints on the properties of the galaxies that hosted them, even when they are too faint for direct detection, providing insight into the nature of photon production at high- z . While this may sound only of niche interest, it is in fact particularly important, since this epoch is the time at which the universe underwent its last major phase change from being predominantly neutral to being mostly ionized. Photons produced through some astrophysical processes at this time will have driven this transition. Indeed, it is most likely that the massive star progenitors of long GRBs are also copious producers of ionizing photons and may well be the kind of star that drove this reionization.

1.11.4 New Classes of GRB

One of the major features of *Swift* in comparison to previous GRB missions is the sheer number of well-localized bursts it detects. This now numbers well over 1000. This detection rate means that in addition to observing many long and short GRBs, it also has the capability to locate rare and hitherto unexpected populations of events. Crucial to this capability is that unlike predecessor missions, the *Swift*-BAT can also reconstruct an image of the gamma-ray sky, in contrast to simple rate detectors, which rely on observing the source as an increase in rate across the detector. Because of this, it is particularly well suited for studying long-lived objects that do not greatly increase the gamma-ray rate but whose integrated fluence can be large. In this regard, *Swift* has been crucial in expanding the range of duration over which high-energy transients can be discovered (see Figure 1.8). Two crucial populations observed by this technique have been the ULGRBs (Levan et al. 2014) and the relativistic tidal disruption events (Bloom et al. 2011; Levan et al. 2011; Burrows et al. 2011).

The ULGRBs are, as the name suggests, GRBs that appear to last much longer than the majority of GRBs. Indeed, the GRB duration distribution largely peters out at durations of around 1000 s. In contrast, the ULGRBs appear to have durations of

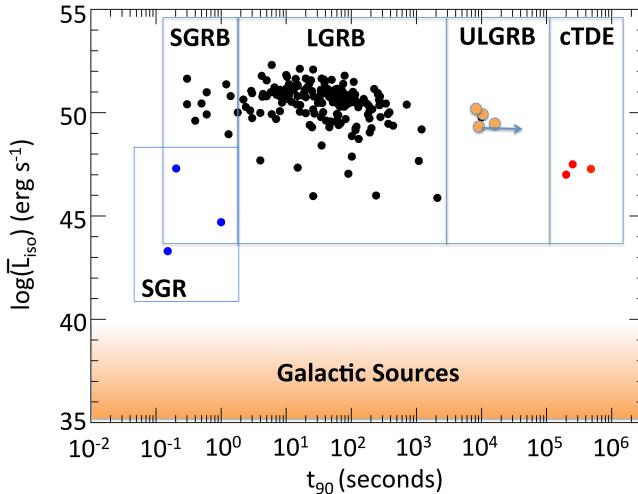


Figure 1.8. Parameter space (in the duration versus absolute luminosity plane) for GRB-like phenomena. The well-known long- and short-duration GRBs are shown, as are populations of galactic soft-gamma repeaters (SGRs) that may appear as short GRBs in external galaxies. New populations of events have been uncovered by *Swift* at extreme durations of >1000 s, largely thanks to its imaging capability. These are the ULGRBs and candidate relativistic tidal disruption events (cTDEs). Figure from Levan et al. (2016). © 2016.

several thousands, in some cases up to 20,000 s (Gendre et al. 2013). This must reflect a central engine that is active for a far longer period than in normal long GRBs. The long duration may be created by longer lived accretion, for example because the infall time of the material to the black hole is longer. This might be the case if the collapsing star has a larger radius, as would be the case if it were a giant (Levan et al. 2014; Gendre et al. 2013), perhaps the creation of a magnetar central engine that lasts for an order of magnitude longer than in other GRBs (Greiner et al. 2015; Metzger et al. 2015) or alternatively could be indicative of an altogether different progenitor, for example, the tidal disruption of a white dwarf by an intermediate-mass black hole (MacLeod et al. 2016). Some observations to date posit a possible link between the ULGRBs and another mysterious population of transients known as the superluminous supernovae (SLSNe), as one GRB appears to be accompanied by a supernova a factor of a few brighter than that seen in other bursts. However, further observations of these rare but extreme events are clearly needed to answer the question of their origins.

The relativistic tidal disruption events are (at least in principle) more readily explained. It has long been expected that massive black holes in galactic nuclei should occasionally disrupt main-sequence or giant stars that stray within their tidal radius, the point at which the gravitational force on the outer edge of a star from the black hole exceeds the gravity of the star (Rees 1988). A number of candidates for such events have been located over the years, identified as being bright nuclear transients with thermal spectra and temperatures of $>20,000$ K. However, *Swift* has located a population in which the thermal emission is not readily visible and where it seems a moderately relativistic jet is also launched. These relativistic tidal disruption

events provide a new route to measuring the ubiquity of black holes in galaxies (particularly in dwarf galaxies), as well as answering questions about black hole growth and particle acceleration.

1.12 New Insights from *Fermi*

In 2008, the *Fermi* gamma-ray satellite was launched. *Fermi* is a purely gamma-ray observatory and does not have the ability to rapidly localize GRB afterglows. However, it is equipped with extremely powerful gamma-ray detectors that have excellent sensitivity and, crucially, a far wider energy range than the *Swift*-BAT. Its gamma-ray burst monitor (GBM) has an effective energy response from ~ 10 keV to ~ 25 MeV, while the Large Area Telescope (LAT) is a pair-conversion detector⁸ that can observe photons with energies of > 300 GeV. This allows significant observations of GRBs in its own right, but combined with the capabilities of *Swift* and ground-based optical-to-radio follow-up, it provides a more complete panchromatic picture of GRB explosions. Observations with *Fermi* have been valuable in probing the nature of the prompt GRB emission and have yielded numerous important results.

Perhaps most importantly, these observations have shown that the GRB spectrum extends to very high energies, with *Fermi*-LAT-detecting individual photons from bursts up to tens of GeV in energy (Abdo et al. 2009a, 2009b). Interestingly, these photons often appear delayed with respect to the rest of the prompt emission, sometimes by thousands of seconds, and imply that there is an additional component to the spectral (and temporal) structures of GRB prompt emission that has been hitherto unseen.

These very high energy photons also provide a route for testing quantum gravity models, in particular, varieties in which the speed of light varies as a function of photon energy. While the very high energy photons seen in some *Fermi* GRBs arrive much later, some arrive at the same time as the prompt emission begins. While the strength of the constraints available on quantum gravity depends strongly on the assumed physics of the emission (in particular, the assumption that photons of different energies are produced nearly simultaneously at the source), the long path length traversed by the gamma-ray photons before detection yields strong constraints on the variation of the speed of light, suggesting that quantum gravity effects must only occur on mass scales significantly above the Planck mass (Abdo et al. 2009a).⁹

Fermi has also provided a much better view of the broadband properties of the prompt γ -ray emission, where its wide energy coverage enables hitherto undetected features to be found. Of particular import here is the discovery that many bursts are not simply described by a power law or a combination of power laws but show a clear thermal component with a temperature of tens to hundreds of keV. This is

⁸A pair-conversion detector is one in which high-energy photons interact with the detector, producing electron–positron pairs. These pairs then pass through the detector as charged particles that are readily detected, providing information about the direction and energy of the incoming photon.

⁹Here the photon energy relates to the Planck mass as $E_\gamma = M_p c^2 = 1.22 \times 10^{19}$ GeV, where M_p is the Planck mass.

often referred to as the photospheric emission and seems to be a common feature in GRBs. Such thermal emission was not necessarily surprising, since an optically thick fireball is naturally thermal in nature, and early work had suggested the possibility of thermal emission. Nonetheless, the power-law nature of the majority of GRBs seen by BATSE instead suggested that nonthermal, most likely synchrotron, processes were at work. The observations with *Fermi* instead suggest a more nuanced picture, in which both thermal and nonthermal components are at play. This, at least in principle, provides a new route to understanding the creation and evolution of the prompt emission.

1.13 Multimessenger Astronomy

A particular appeal of studies of GRBs is that their extreme progenitor systems are the most natural sites for the creation of additional astrophysical messengers. These “nonphotonic” signals include observations from very high energy cosmic rays, neutrinos, and gravitational waves.

1.13.1 Gravitational Waves

Gravitational waves naturally arise from accelerating nonspherically symmetric systems, so merging neutron stars or asymmetries during the core-collapse process are promising routes to their generation. The watershed detection of gravitational waves from a merging double black hole system in 2015 September (Abbott et al. 2016) provided a crucial demonstration that gravitational-wave interferometers had reached the necessary sensitivity to yield direct detections, measuring dimensionless strains (the change in path length due to the passage of the wave divided by the total path length, $\Delta L/L$) of order $\sim 10^{-22}$. These initial detections of black hole–black hole mergers did not yield any clear electromagnetic counterparts, and this is largely to be expected. However, a controversial possible detection of a short GRB was made by *Fermi* (Connaughton et al. 2016; Greiner et al. 2016), and such signals can be created in rather more exotic models of black hole–black hole formation and merger.

The crucial breakthrough for multimessenger astronomy occurred in 2017 August, when *Fermi* detected a short-duration GRB approximately 2 s after LIGO and VIRGO observed the first merger of a binary neutron star system (Abbott et al. 2017b; Figure 1.9). Since short GRBs are rare, the connection between the two based on timing alone was highly suggestive. This became increasingly clear as location information became available. The position triangulated by the two LIGO detectors and VIRGO was in agreement with that from *Fermi* and with a refined electromagnetic position enabled by an independent detection of the GRB by the *INTEGRAL* satellite. The most accurate position enabled by the gravitational-wave detectors spanned only $\sim 30 \text{ deg}^2$, and the gravitational waves from the merger itself also enabled a distance of approximately 40 Mpc to be determined. This region contained only a few tens of galaxies cataloged at the appropriate distance, so it could be searched relatively efficiently for a counterpart at longer wavelengths. Indeed, while the number of galaxies to survey was a significant observational challenge, it was made more difficult still by the sky location of the source, which

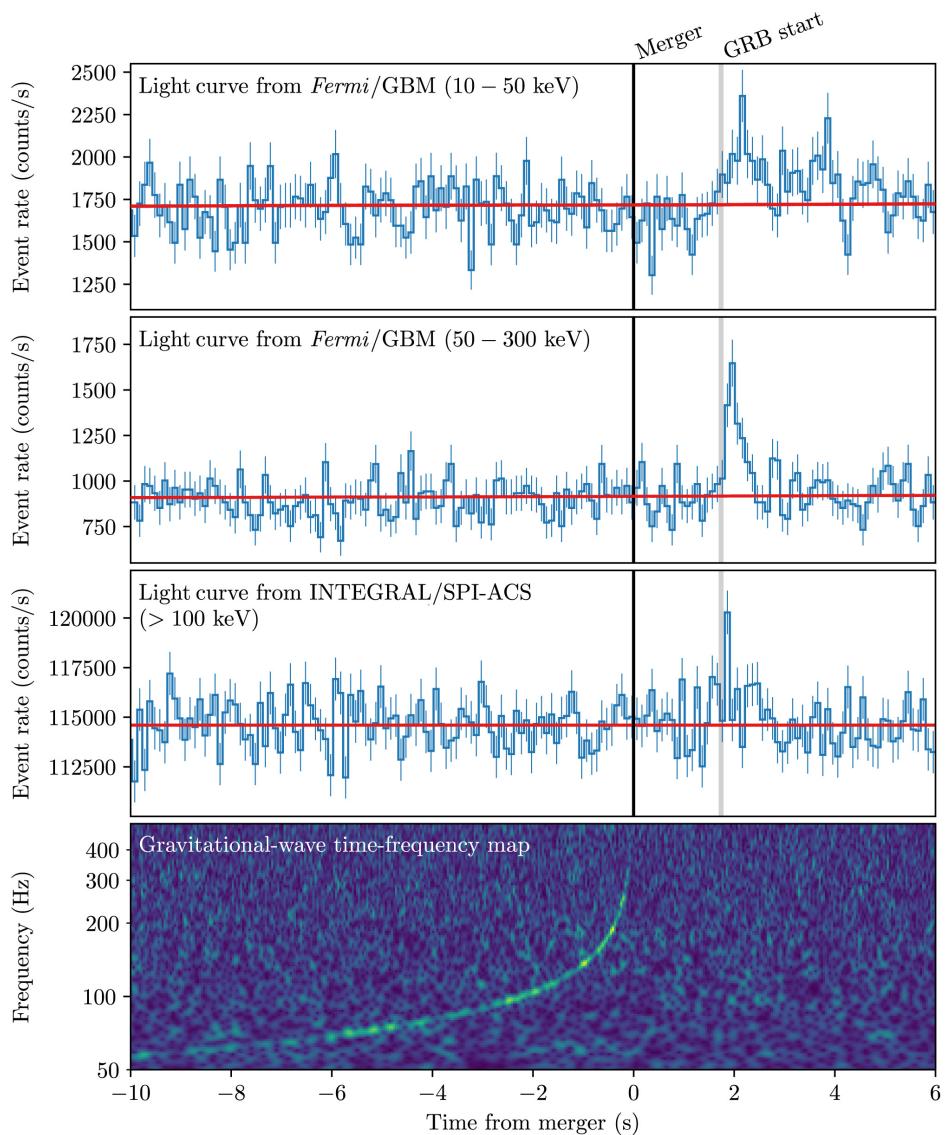


Figure 1.9. Detection of GW 170817 and GRB 170817A. The lower panel shows the frequency of the gravitational-wave signal plotted against the time of the merger, with the intensity demonstrating the strength of the signal. The sweeping line is often referred to as the gravitational-wave chirp, when the source becomes “louder” and at higher frequencies as the orbit decays toward merger. The upper panels show the gamma-ray light curves as recorded by *Fermi* (in two different energy bands) and *INTEGRAL*. The GRB occurred within ~ 2 s of the merger. Given the typical rate of short GRBs, this is already strongly indicative of an association, even before the addition of positional information or the identification of a counterpart at other wavelengths.

© 2017. The American Astronomical Society. All rights reserved. CC.BY.3.0

was only $\sim 50^\circ$ from the Sun, and thus only visible to observers in the hour after twilight.

Nonetheless, observations targeted at galaxies, as well as the wider error region, were undertaken by several groups from Chile on the evening of August 17. Several groups included in their observations the galaxy NGC 4993 (Figure 1.10), which clearly contained a new source, first identified by the Swope telescope (Coulter et al. 2017) and subsequently independently confirmed by several other observers (Arcavi et al. 2017; Drout et al. 2017; Soares-Santos et al. 2017; Tanvir et al. 2017; Lipunov et al. 2017). This new source was not there several weeks earlier and was at the right distance for the gravitational-wave counterpart. Subsequent observations revealed the source to have extreme and unprecedented photometric and spectroscopic evolution. It evolved rapidly from blue to red and showed a series of never-before-observed spectral features. These observations marked it as a kilonova

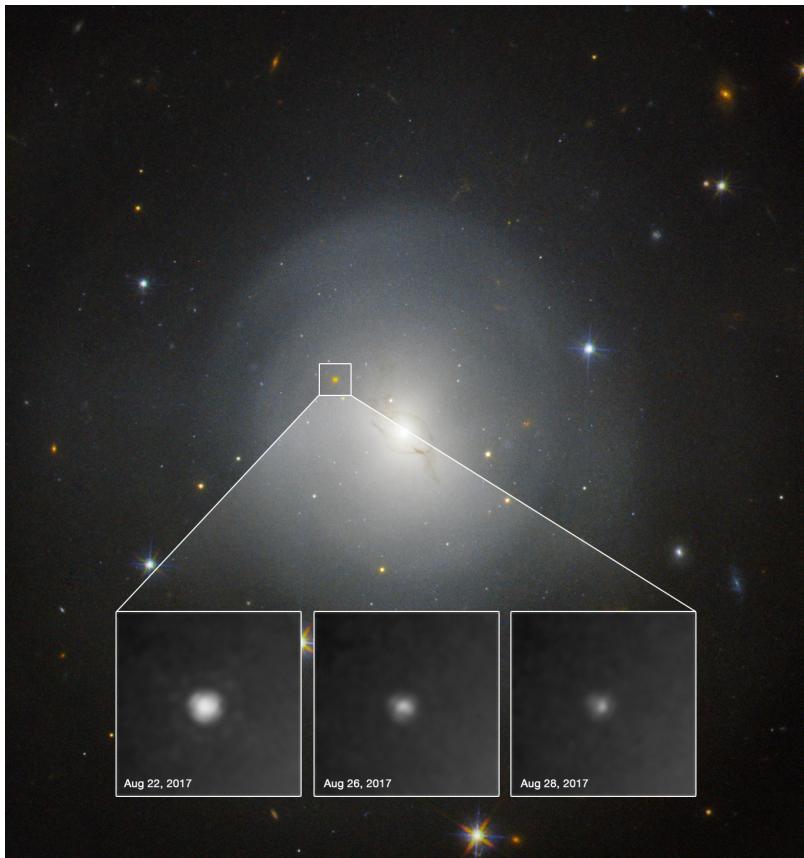


Figure 1.10. NGC 4993 and the kilonova AT 2017gfo, associated with GW 170817 and GRB 170817A, as viewed with the *HST*. The host galaxy is bright and highly extended. This is largely because at a distance of ~ 40 Mpc, this kilonova is located a factor of ~ 40 closer than the kilonova in GRB 130603B (see Figure 1.6). The kilonova is also very bright and can be seen to fade over the course of several days between the observations. Courtesy of NASA.

(Pian et al. 2017; Shappee et al. 2017; Cowperthwaite et al. 2017; Smartt et al. 2017; Tanvir et al. 2017) and conclusively demonstrated its association with the gravitational-wave source. It is widely interpreted that the early blue emission (often referred to as the blue kilonova) is caused by material in which “light” *r*-process elements are being produced and there is a modest electron fraction. In contrast, the red component (the red kilonova) is generally thought to arise from material with a much lower electron fraction that is synthesizing heavier elements, such as the lanthanides. It is much redder because of the high opacity of lanthanide material.

These pivotal observations finally opened the opportunity of multimessenger astronomy using gravitational waves and visible light as a route to probe the final moments and aftermath of the merger of a compact binary. The kilonova was clearly powered by the synthesis of heavy *r*-process elements, although the elemental makeup remains uncertain. The detection also opens the possibility new tests of modified theories of gravity and an alternative route to measurements of the Hubble constant (Abbott et al. 2017a). At the time of writing, intensive observations were ongoing that had detected the evolution of the off-axis GRB afterglow emission across the electromagnetic spectrum (Hallinan et al. 2017; Troja et al. 2017; Lyman et al. 2018) and should ultimately reveal the details of the jet geometry and probe the similarity of the GRB associated with GW 170817 with the majority of short GRBs.

1.13.2 High-energy Neutrinos

Once the GRB is launched, particle acceleration in the GRB jets clearly accelerates many electrons to relativistic velocities. This particle acceleration could then be a promising site for the acceleration of the most energetic cosmic rays but would require the acceleration of more massive particles than electrons (e.g., protons or heavier nuclei). The majority of ultra-high-energy cosmic rays are protons, but they cannot be readily identified with a given GRB directly; they are massive and charged particles whose trajectories are altered by magnetic fields within galaxies and cannot travel at light speed. Because of this, we will not observe a burst of cosmic rays in association with any GRB. However, the particle acceleration processes that drive proton acceleration should also produce a shower of neutrinos because the protons will interact with gamma rays to produce pions, and these pions will decay to muons and then positrons, with the decay chain releasing three neutrinos (Waxman 1995). These neutrinos are typically high-energy, with energies in the TeV range, and thus are readily identifiable from the lower-energy (MeV) supernova or solar neutrinos. It is therefore the observation of neutrinos from GRBs that would provide evidence for proton acceleration in their jets. However, while large neutrino detectors, such as IceCube, have now reached the necessary sensitivity to observe such signatures, none has yet been found.

1.14 Summary

The study of GRBs has come a long way from the detection of the first high-energy photons over 60 yr ago. Many, although by no means all, of the mysteries surrounding the nature of GRBs have given way to intensive study, led by both

increasingly impressive physical insight and dramatically improved technology. Although the multiwavelength studies of GRBs and their host galaxies continue to be intensive exercises, the strategies and facilities that are now in place allow the observations to be undertaken in a relatively routine manner, at least most of the time.

As progress has been made toward GRB origins, their value as probes of the universe has become increasingly clear. They are useful across the full gambit of traditional astrophysical disciplines. Immediately before their collapse, they are representative of extreme conditions in stellar and binary evolution and thus can test our understanding of it. The bursting phase involves the creation of black holes, can create powerful gravitational waves, and accelerates material to relativistic velocity, providing a unique testbed for extreme physics. As the shock from the burst propagates outward, it interacts with the surrounding circumstellar and interstellar medium, enabling us to determine the environments around massive stars across the universe. We can do this because the outgoing material forms shocks, enabling us to probe shock physics at extreme velocities. These shocks set up such luminous afterglows that we can see GRBs across the universe—we can probe the distant universe directly with the bursts. If this did not make them sufficiently powerful cosmological probes, the fact that they are related to massive stars means that we can use them to trace star formation rates, while the metallicities derived from their afterglows mean that we can track the buildup of both stellar mass and metallicity across cosmic time.

Although there are no doubt numerous fields of astrophysics where results derived from GRBs have not had significant impact, their study has made pivotal contributions across many of the most-studied fields in contemporary astrophysics. The aim of this book is to present many of these most important contributions and place them in context. The following chapters are broken down predominantly into different regions of GRB study and may appeal to different audiences. They are intended (to the extent possible) to be self-contained, so there is no requirement to read them consecutively, or even to read all of them, if there is one particular aspect of GRB study that is of interest.

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, *PhRvL*, 116, 061102
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *Natur*, 551, 85
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, *ApJL*, 848, L12
- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2009a, *Natur*, 462, 331
- Abdo, A. A., Ackermann, M., Arimoto, M., et al. 2009b, *Science*, 323, 1688
- Akerlof, C., Balsano, R., Barthelmy, S., et al. 1999, *Natur*, 398, 400
- Arcavi, I., Hosseinzadeh, G., Howell, D. A., et al. 2017, *Natur*, 551, 64
- Arzoumanian, Z., Chernoff, D. F., & Cordes, J. M. 2002, *ApJ*, 568, 289
- Barnes, J., & Kasen, D. 2013, *ApJ*, 775, 18
- Berger, E. 2010, *ApJ*, 722, 1946
- Berger, E., Fong, W., & Chornock, R. 2013, *ApJL*, 774, L23
- Bloom, J. S., Djorgovski, S. G., & Kulkarni, S. R. 2001, *ApJ*, 554, 678

- Bloom, J. S., Frail, D. A., & Kulkarni, S. R. 2003, *ApJ*, 594, 674
- Bloom, J. S., Giannios, D., Metzger, B. D., et al. 2011, *Science*, 333, 203
- Bloom, J. S., Prochaska, J. X., Pooley, D., et al. 2006, *ApJ*, 638, 354
- Brecher, K., & Morrison, P. 1974, *ApJL*, 187, L97
- Briggs, M. S., Paciesas, W. S., Pendleton, G. N., et al. 1996, *ApJ*, 459, 40
- Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2013, *ApJ*, 764, 179
- Bromm, V., & Loeb, A. 2006, *ApJ*, 642, 382
- Burrows, D. N., Kennea, J. A., Ghisellini, G., et al. 2011, *Natur*, 476, 421
- Caraveo, P. A., Mignani, R. P., Tavani, M., & Bignami, G. F. 1997, *A&A*, 326, L13
- Church, R. P., Levan, A. J., Davies, M. B., & Tanvir, N. 2011, *MNRAS*, 413, 2004
- Connaughton, V., Burns, E., Goldstein, A., et al. 2016, *ApJL*, 826, L6
- Costa, E., Frontera, F., Heise, J., et al. 1997, *Natur*, 387, 783
- Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, *Sci*, 358, 1556
- Cowperthwaite, P. S., Berger, E., Villar, V. A., et al. 2017, *ApJL*, 848, L17
- Cucchiara, A., Levan, A. J., Fox, D. B., et al. 2011, *ApJ*, 736, 7
- D'Avanzo, P. 2015, *JHEAp*, 7, 73
- Della Valle, M., Chincarini, G., Panagia, N., et al. 2006, *Natur*, 444, 1050
- Drout, M. R., Piro, A. L., Shappee, B. J., et al. 2017, *Sci*, 358, 1570
- Duncan, R. C., & Thompson, C. 1992, *ApJL*, 392, L9
- Evans, P. A., Beardmore, A. P., Page, K. L., et al. 2009, *MNRAS*, 397, 1177
- Fong, W., & Berger, E. 2013, *ApJ*, 776, 18
- Fong, W., Berger, E., Chornock, R., et al. 2013, *ApJ*, 769, 56
- Fong, W., Berger, E., Margutti, R., & Zauderer, B. A. 2015, *ApJ*, 815, 102
- Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M., & Taylor, G. B. 1997, *Natur*, 389, 261
- Frail, D. A., Kulkarni, S. R., Sari, R., et al. 2001, *ApJL*, 562, L55
- Fruchter, A. S., Levan, A. J., Strolger, L., et al. 2006, *Natur*, 441, 463
- Fynbo, J. P. U., Jakobsson, P., Prochaska, J. X., et al. 2009, *ApJS*, 185, 526
- Fynbo, J. P. U., Watson, D., Thöne, C. C., et al. 2006, *Natur*, 444, 1047
- Gal-Yam, A., Fox, D. B., Price, P. A., et al. 2006, *Natur*, 444, 1053
- Galama, T. J., Vreeswijk, P. M., van Paradijs, J., et al. 1998, *Natur*, 395, 670
- Gehrels, N., Barthelmy, S. D., Burrows, D. N., et al. 2008, *ApJ*, 689, 1161
- Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, *ApJ*, 611, 1005
- Gehrels, N., Norris, J. P., Barthelmy, S. D., et al. 2006, *Natur*, 444, 1044
- Gehrels, N., Sarazin, C. L., O'Brien, P. T., et al. 2005, *Natur*, 437, 851
- Gendre, B., Stratta, G., Atteia, J. L., et al. 2013, *ApJ*, 766, 30
- Giacconi, R., Gursky, H., Paolini, F. R., & Rossi, B. B. 1962, *PhRvL*, 9, 439
- Greiner, J., Burgess, J. M., Savchenko, V., & Yu, H.-F. 2016, *ApJL*, 827, L38
- Greiner, J., Mazzali, P. A., Kann, D. A., et al. 2015, *Natur*, 523, 189
- Greiner, J., Yu, H.-F., Krühler, T., et al. 2014, *A&A*, 568, A75
- Haislip, J. B., Nysewander, M. C., Reichart, D. E., et al. 2006, *Natur*, 440, 181
- Hallinan, G., Corsi, A., Mooley, K. P., et al. 2017, *Sci*, 358, 1579
- Harwit, M., & Salpeter, E. E. 1973, *ApJL*, 186, L37
- Hjorth, J., Malesani, D., Jakobsson, P., et al. 2012, *ApJ*, 756, 187
- Hjorth, J., Sollerman, J., Møller, P., et al. 2003, *Natur*, 423, 847
- Hjorth, J., Watson, D., Fynbo, J. P. U., et al. 2005, *Natur*, 437, 859
- Hurley, K., Boggs, S. E., Smith, D. M., et al. 2005, *Natur*, 434, 1098

- Jin, Z.-P., Hotokezaka, K., Li, X., et al. 2016, *NatCo*, **7**, 12898
- Jin, Z.-P., Li, X., Cano, Z., et al. 2015, *ApJL*, **811**, L22
- Kawai, N., Kosugi, G., Aoki, K., et al. 2006, *Natur*, **440**, 184
- Klebesadel, R., Evans, W. D., Laros, J. G., et al. 1982, *ApJL*, **259**, L51
- Klebesadel, R. W., Strong, I. B., & Olson, R. A. 1973, *ApJL*, **182**, L85
- Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, *ApJL*, **413**, L101
- Krimm, H. A., Vanderspek, R. K., & Ricker, G. R. 1994, in AIP Conf. Proc. 307, Searches for optical counterparts of BATSE gamma-ray bursts, ed. G. J. Fishman (Melville, NY: AIP), 423
- Lamb, D. Q. 1995, *PASP*, **107**, 1152
- Lamb, D. Q., Lamb, F. K., & Pines, D. 1973, *NPhS*, **246**, 52
- Levan, A., Crowther, P., de Grijs, R., et al. 2016, *SSRv*, **202**, 33
- Levan, A. J., Tanvir, N. R., Cenko, S. B., et al. 2011, *Sci*, **333**, 199
- Levan, A. J., Tanvir, N. R., Starling, R. L. C., et al. 2014, *ApJ*, **781**, 13
- Li, L.-X., & Paczyński, B. 1998, *ApJL*, **507**, L59
- Lipunov, V. M., Gorbovskoy, E., Kornilov, V. G., et al. 2017, *ApJL*, **850**, L1
- Littlejohns, O. M., Tanvir, N. R., Willingale, R., et al. 2013, *MNRAS*, **436**, 3640
- Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, *Sci*, **318**, 777
- Lyman, J. D., Lamb, G. P., Levan, A. J., et al. 2018, *NatAs*, **2**, 751
- MacLeod, M., Guillotin-Plantier, J., Ramirez-Ruiz, E., Kasen, D., & Rosswog, S. 2016, *ApJ*, **819**, 3
- Mazets, E. P., & Golenetskii, S. V. 1981, *Ap&SS*, **75**, 47
- Mazets, E. P., Golenetskii, S. V., & Il'Inskii, V. N. 1974, *ZhPmR*, **19**, 126
- Metzger, B. D., Berger, E., & Margalit, B. 2017, *ApJ*, **841**, 14
- Metzger, B. D., Margalit, B., Kasen, D., & Quataert, E. 2015, *MNRAS*, **454**, 3311
- Metzger, M. R., Djorgovski, S. G., Kulkarni, S. R., et al. 1997, *Natur*, **387**, 878
- Mukherjee, S., Feigelson, E. D., Jogesh Babu, G., et al. 1998, *ApJ*, **508**, 314
- Nakar, E., & Sari, R. 2012, *ApJ*, **747**, 88
- Nemiroff, R. J. 1994, *ComAp*, **17**, 189
- Nysewander, M., Fruchter, A. S., & Pe'er, A. 2009, *ApJ*, **701**, 824
- Paciesas, W. S., Meegan, C. A., Pendleton, G. N., et al. 1999, *ApJs*, **122**, 465
- Paczynski, B. 1995, *PASP*, **107**, 1167
- Perley, D. A., Krühler, T., Schulze, S., et al. 2016, *ApJ*, **817**, 7
- Pian, E., D'Avanzo, P., Benetti, S., et al. 2017, *Natur*, **551**, 67
- Piro, L., Troja, E., Gendre, B., et al. 2014, *ApJL*, **790**, L15
- Racusin, J. L., Karpov, S. V., Sokolowski, M., et al. 2008, *Natur*, **455**, 183
- Rees, M. J. 1988, *Natur*, **333**, 523
- Rees, M. J. 1995, *PASP*, **107**, 1176
- Rees, M. J., & Meszaros, P. 1994, *ApJL*, **430**, L93
- Sahu, K. C., Livio, M., Petro, L., et al. 1997, *Natur*, **387**, 476
- Salvaterra, R., Della Valle, M., Campana, S., et al. 2009, *Natur*, **461**, 1258
- Savaglio, S., Glazebrook, K., & Le Borgne, D. 2009, *ApJ*, **691**, 182
- Shappee, B. J., Simon, J. D., Drout, M. R., et al. 2017, *Sci*, **358**, 1574
- Smartt, S. J., Chen, T.-W., Jerkstrand, A., et al. 2017, *Natur*, **551**, 75
- Soares-Santos, M., Holz, D. E., Annis, J., et al. 2017, *ApJL*, **848**, L16
- Stanek, K. Z., Matheson, T., Garnavich, P. M., et al. 2003, *ApJL*, **591**, L17
- Strong, I. B., Klebesadel, R. W., & Olson, R. A. 1974, *ApJL*, **188**, L1

- Svensson, K. M., Levan, A. J., Tanvir, N. R., Fruchter, A. S., & Strolger, L.-G. 2010, *MNRAS*, **405**, 57
- Tanvir, N. R., Fox, D. B., Levan, A. J., et al. 2009, *Natur*, **461**, 1254
- Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. 2013, *Natur*, **500**, 547
- Tanvir, N. R., Levan, A. J., González-Fernández, C., et al. 2017, *ApJL*, **848**, L27
- Thornton, D., Stappers, B., Bailes, M., et al. 2013, *Sci*, **341**, 53
- Toma, K., Sakamoto, T., & Mészáros, P. 2011, *ApJ*, **731**, 127
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, **613**, 898
- Troja, E., Piro, L., van Eerten, H., et al. 2017, *Natur*, **551**, 71
- Tunnicliffe, R. L., Levan, A. J., Tanvir, N. R., et al. 2014, *MNRAS*, **437**, 1495
- van Paradijs, J., Groot, P. J., Galama, T., et al. 1997, *Natur*, **386**, 686
- Waxman, E. 1995, *PhRvL*, **75**, 386
- Wheaton, W. A., Ulmer, M. P., Baity, W. A., et al. 1973, *ApJL*, **185**, L57
- Woosley, S. E. 1993, *ApJ*, **405**, 273

Gamma-Ray Bursts

Andrew Levan

Chapter 2

Prompt Emission

2.1 Observational Properties

The brief flash of gamma rays emitted in the creation of a GRB was the route by which GRBs were detected and the property for which they are named. The prompt emission is readily detected, even by rudimentary space-based gamma-ray detectors, due to the extreme high-energy photon budget that GRBs exhibit. Indeed, at peak, GRBs outshine all other sources within the gamma-ray sky, including the Sun. Since GRBs have been detected in gamma rays for more than 50 yr, much of their gamma-ray emission has now been mapped out. We know a great deal about their temporal behavior, the spectra from the soft X-ray to very hard gamma-ray, and the apparent distributions of key properties (peak flux, fluence, peak energy, etc). However, while the observational properties are usually well measured, there remain several crucial questions relating to their interpretation and the origins of several of the features seen.

2.1.1 Temporal Structure

There is an often-spoken adage in the community about GRBs that states “when you’ve seen one GRB, you’ve seen one GRB.” This reflects that GRB prompt emission is hugely variable burst to burst. There are no two bursts with the same duration and light curves that look identical. Some bursts are smooth, others highly erratic; some last for a fraction of a second, others for several hours. The prompt gamma-ray light curves of several well-known GRBs are shown in Figure 2.1. These demonstrate the range of behavior that is commonly seen. This includes highly variable spiky bursts and those that exhibit a fast rise and exponential decay, a so-called FRED profile. It may seem that if GRBs are so diverse, they should have many different origins. However, closer examination of GRB prompt emission suggests that while variation from burst to burst is clear, there are also many unifying features, and indeed, bursts can be categorized according to these.

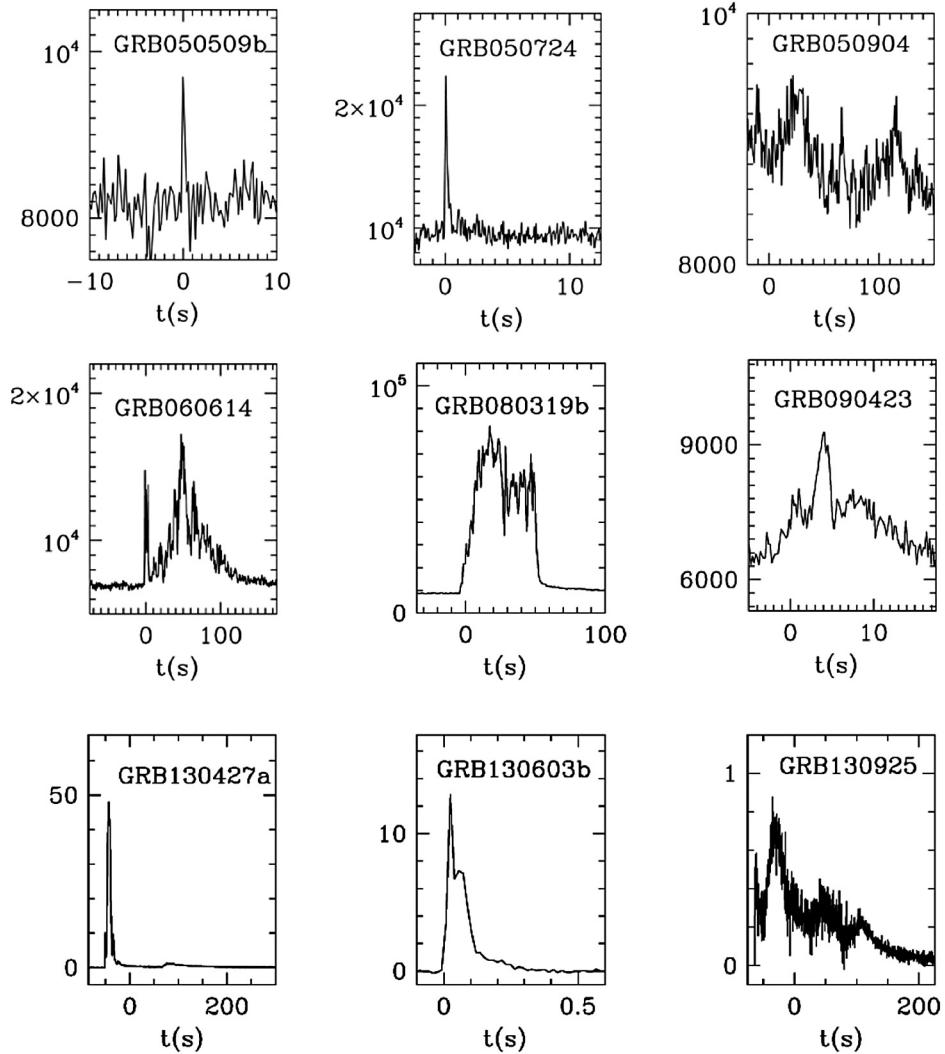


Figure 2.1. Prompt gamma-ray light curves from the *Swift*-BAT for a set of well-studied *Swift* bursts. The wide range of durations and burst morphologies can be clearly seen, highlighting the diversity of GRB prompt emission.

Perhaps the most natural property to consider is the duration of the burst, measuring how long the gamma rays are visible. The most commonly used measure of duration is the so-called t_{90} . This is the period of time over which 90% of the observed fluence (or energy) is observed. It can be calculated by plotting the cumulative distribution of flux with time and determining the period between 5% and 95% within this. Similar diagnostics, such as the narrower t_{50} , which encompasses 50% of the total energy, can also be used. In practice, it is often the integral of the counts that is used to determine the duration, which may not precisely match the flux if the spectrum of the emission is variable throughout the burst (see Section 2.1.2 for more information).

The advantage of a duration measurement is that it is independent of the morphology of the light curve. Hence, it allows highly variable bursts to be compared to those with much smoother light curves regarding the duration over which the prompt emission is active. Duration measurements were used to investigate possible distinctions between populations of GRBs as early as the 1980s (Mazets & Golenetskii 1981). From the measured durations, it became apparent that there are at least two broad populations of GRBs. So-called short bursts have a typical duration of ~ 1 s, and long bursts have an average duration of about 30 s (Mazets & Golenetskii 1981; Kouveliotou et al. 1993). The distribution of durations in both populations is reasonably modeled by a lognormal distribution, with a dividing line between the two groups customarily drawn at $t_{90} = 2$ s. However, it is clearly the case that the distributions of both groups continue beyond this; short bursts can have durations $t_{90} > 2$ s, and long burst can have $t_{90} < 2$ s.

Examination of the spectral properties of the different populations (see Section 2.1.2) suggests that short bursts emit more high-energy radiation, while long bursts emit somewhat less. This is often encapsulated in a so-called hardness ratio, which reflects either the count rate or energy contained within one band compared to another. For example, for bursts observed with BATSE, it is common to define the hardness ratio as

$$HR_{32} = \frac{S_3(100 - 300 \text{ keV})}{S_2(50 - 100 \text{ keV})}, \quad (2.1)$$

where S_3 and S_2 are the fluence of the burst in the 100–300 and 50–100 keV range. There is also an S_1 that covers the 25–50 keV range, and indeed, hardness ratios can be defined in any arbitrary band. Since it is common to express the hardness ratios logarithmically, they can be thought of as comparable to colors used in magnitude systems in the optical and IR. In this analogy, a hard burst is the equivalent of a blue source.

The addition of spectral hardness as a parameter increases the distinction between the long- and short-burst populations, with short bursts appearing spectrally harder (i.e., exhibiting more higher-energy photons) than the long bursts. For this reason, it is also common to see the populations referred to as short-hard and long-soft bursts. The hardness-duration diagram for GRBs is shown in Figure 2.2.

It is also apparent that the details of the bursts depend on the energy range viewed. There is evidence that in the long bursts, the higher-energy emission arrives earlier than the lower-energy emission, leading to an evolution from hard to soft. This so-called spectral lag is generally positive for long bursts and normally zero for short bursts (e.g., Gehrels et al. 2006). It is also clear that bursts may appear to last for a longer duration at lower energies, perhaps because of the lag, the relative sensitivity of the detector, or even a contribution from the early afterglow. This means that the duration of a burst as recorded is a function of the energy range and sensitivity of the detector used to make the observations. These differing durations imply that the distinction between long and short GRBs is not as clean as initially suspected and instead is a function of the detector used to make the observation.

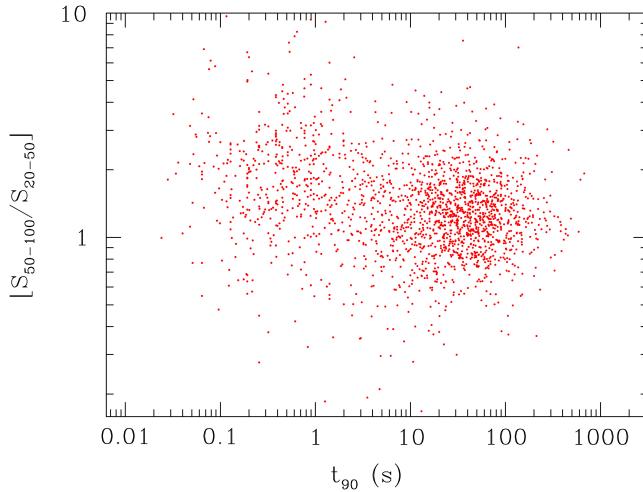


Figure 2.2. Hardness-duration diagram for BATSE GRBs (based on the BATSE 4B catalog and updated from Kouveliotou et al. 1993), demonstrating the divide between short-hard bursts ($t_{90} < 2$ s) and long-short bursts ($t_{90} > 2$ s). It is apparent that the median hardness of the short bursts is significantly higher than for the long-burst population, so a combination of hardness and duration provides a better discriminant of the nature of a given burst.

Indeed, while it appears that 2 s is a good divide for the BATSE population, it has been suggested that for the *Swift*-BAT, whose soft response extends to ~ 15 keV, the duration at which there is a 50% chance of lying in either population is in fact $t_{90} \sim 0.7$ s (Bromberg et al. 2013).

These observational constraints complicate attempts to place GRBs with durations of $\sim 1\text{--}5$ s firmly within one class or another and lead to the requirement of using additional diagnostics, often from the afterglow or host galaxy observations, to determine the actual class of any event (Levan et al. 2007; Zhang et al. 2009). Indeed, this is further complicated, since suggestions persist of a possible third group of GRBs with soft spectra and intermediate duration (Mukherjee et al. 1998). While the existence of such a group as a distinct physical phenomenon has not been confirmed, it further complicates classifications and studies that are based purely on the prompt duration of a given GRB. More recently, *Swift* has begun to discover events with far longer durations, so-called ultra-long GRBs (Gendre et al. 2013; Levan et al. 2014). These bursts have been argued to be both the long tail of the long GRBs (Zhang et al. 2014) and a unique new population (Boér et al. 2015; Levan 2015).

2.1.2 Spectral Structure

The GRBs also exhibit a wide range of spectral structure that becomes increasingly apparent as the range of energies considered is increased. The emission is frequently nonthermal, or at least contains a nonthermal component. Many observations (for example, those with *Swift*) are adequately fit with simple power-law prescriptions, which dominate over large regions. For the relatively wide bandpass used by BATSE (~ 25 keV to > 1 MeV), a frequently found best-fit model consists of two

smoothly joined power laws with an exponential cutoff, often described as a Band function (after the author who first described it; Band et al. 1993). This function is described (in energy space) as

$$F_E = \begin{cases} A \left(\frac{E}{100 \text{ keV}} \right)^\alpha \exp(-E(2 + \alpha)/E_{\text{peak}}) & \text{if } E < \frac{(\alpha - \beta)E_{\text{peak}}}{(2 + \alpha)} \equiv E_{\text{break}} \\ A \left[\frac{(\alpha - \beta)E_{\text{peak}}}{100 \text{ keV}(2 + \alpha)} \right] \exp((\beta - \alpha)(E/100\text{keV})^\beta) & \text{if } E \geq \frac{(\alpha - \beta)E_{\text{peak}}}{(2 + \alpha)} \end{cases}, \quad (2.2)$$

which, in practice, smoothly joins a low-energy (α) and high-energy (β) spectral slope. The break in the two slopes occurs at $E_{\text{break}} = (\alpha - \beta)E_{\text{peak}}/(2 + \alpha)$, but a more commonly used value is the E_{peak} , which defines the peak of the $\nu F_\nu = EF_E = E^2 N_E$ spectrum (where N_E is the number of photons). This is the photon energy at the peak of the spectrum where the most energy from the burst is emitted. An example of such a spectrum from the famous burst GRB 990123, which reached a peak visual magnitude of ~ 9 , is shown in Figure 2.3. This demonstrates the broken power-law appearance and the clear peak energy. It also shows that over wide ranges of energy, the source will be observed as a single power law. This explains why simple power laws often provide very good representations of the observations from the *Swift*-BAT, since the 15–350 keV range often does not contain the peak energy or have a sufficient lever arm around it to resolve the smooth break between the two regimes.

However, while this spectral function generally provided a good description of the observed spectrum observed by BATSE, it suffered because it offered very limited physical insight, since it was empirically fit to the data and did not provide direct information about the physical processes that produced it. More recent observations enable far greater sensitivity and a much broader energy band. These observations have yielded several new insights into the nature of the prompt phase.

One crucial discovery has been enabled by the launch of the *Fermi* satellite. The improvements in the detectors on board *Fermi* mean that higher-quality data can be obtained on a larger number of bursts. This is important, for example, in temporally resolving the GRB spectra, so that the evolution of the GRB spectral properties with time becomes clear. This is something that was only possible for the brightest bursts observed by BATSE. In addition to a more traditional GRB monitor (GBM) operating in the energy range from a few keV to 30 MeV, it also carries the Large Area Telescope, capable of detecting photons with energies as high as ~ 300 GeV. It is this wide range of energy that has enabled perhaps the most striking results. In particular, it is now apparent that some GRBs do exhibit extremely high-energy emission (see Figure 2.4). This is not naturally expected from the Band model alone, since it predicts that the energy above the peak should cut off rapidly. Instead, the

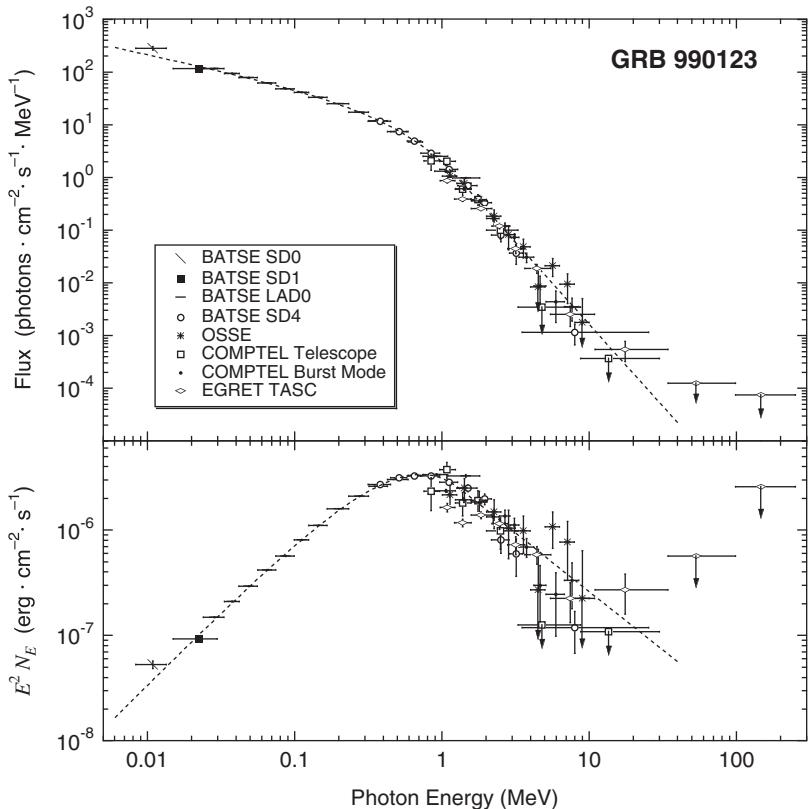


Figure 2.3. The BATSE spectrum of the famous GRB 990123 (Briggs et al. 1999). The burst shows an excellent example of the band spectrum characterized by a low-energy slope of $\alpha = -0.6$ and $\beta = -3.1$, with a peak energy of $E_{\text{peak}} = 720 \text{ keV}$. This can clearly be seen in the lower panel, which plots the emitted energy in the $\nu F_\nu = EF_E = E^2 N_E$ spectrum. Instruments such as the *Swift*-BAT (energy range 15–350 keV) often only sample one of the power-law regions. © 1999. The American Astronomical Society. All rights reserved. Printed in USA.

very high energy emission would appear to form a separate component (Ackerman et al. 2010). This concept is further strengthened, since the very high energy emission is frequently observed well after the bulk of the prompt emission, directly implying that it cannot be tracking the same physical process. It is possible that this is due to inverse Compton scattering of initially much lower energy seed photons.

The second spectral feature that has come to prominence in recent years is that in addition to the nonthermal (frequently power-law) spectra that are observed for many BATSE bursts, a significant fraction of bursts also exhibit a thermal component (Ryde 2005; Ryde et al. 2010). This thermal component typically has a temperature in the region of tens to hundreds of keV (Ryde 2005). In some cases, the burst appears to be dominated by this thermal energy, while in others, it forms only a small contribution. Nonetheless, the presence of a signal that can undoubtedly be well fit by a thermal spectrum provides strong indications about at least some of the emission mechanisms that contribute to the observed prompt emission.

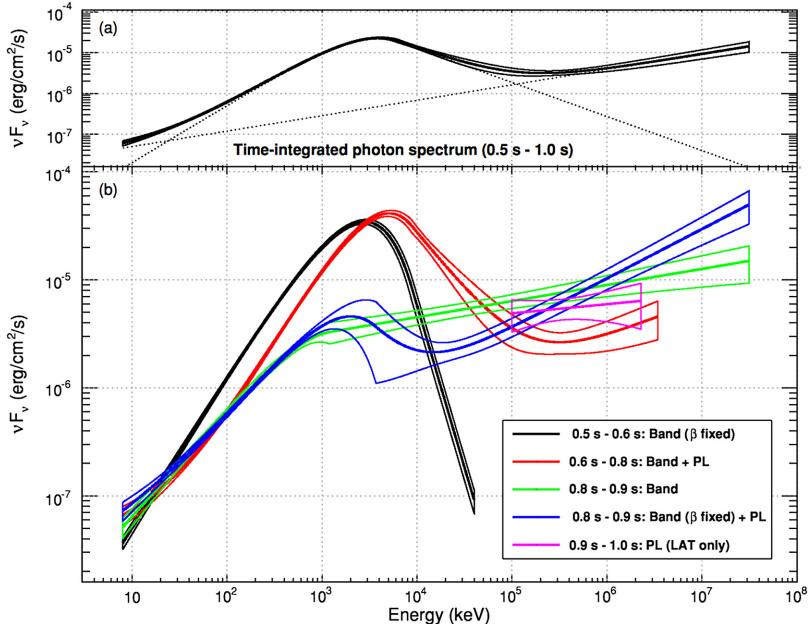


Figure 2.4. Spectrum of the short GRB 090510, as observed by the *Fermi* GBM and LAT instruments (from Ackermann et al. 2010). The band function at energies of hundreds of keV can clearly be seen, but the burst is notable for a further upturn at very high GeV energies, inconsistent with the extrapolation of a band function. © 2010. The American Astronomical Society. All rights reserved.

2.1.3 Polarization

Perhaps one of the most remarkable observations made of GRB prompt emission has been indications of extremely high polarization of the prompt emission. Polarization measurements of gamma rays are far from trivial, and most rely on the dependence of the Compton-scattering cross section (in the detector) on polarization angle. In particular, the measured angular distribution of counts depends on $1 - Q\Pi \cos 2(\phi - \eta)$, where Q is the polarimetric modulation factor and is dependent on the incident photon energy, Π is the degree of polarization, ϕ is the scattering angle, and η is the degree of polarization (Lei et al. 1997; McGlynn et al. 2009). Hence, one obtains a maximum when ϕ and η differ by $\pi/2$ and a minimum when $\phi = \eta$. In principle, the amplitude of the sinusoidal signal then provides a route to measuring the degree of gamma-ray polarization.

However, such measurements are highly challenging and potentially complicated by significant systematic effects and thus mean that any analysis of the prompt gamma-ray polarization of a GRB will be accompanied by a significant associated error. Nonetheless, there have been attempts to make such measurements, first via instruments in which the detector layout fortuitously enabled gamma-ray polarization measurements, and more recently through small but dedicated experiments. All of these observations provide a consistent picture: the prompt emission of GRBs is significantly polarized. The first detections of high polarization (Coburn & Boggs

2003) were subject to significant discussion (Rutledge & Fox 2004), in part because the observations of GRB 021206 indicated a polarization of $P = 80\% \pm 20\%$ (Coburn & Boggs 2003). However, more recent observations support this picture and indicate that the polarization of the prompt GRB emission is indeed frequently at the tens of percent level (Götz et al. 2009; Chattopadhyay et al. 2017). Indeed, in one case, optical polarization taken at early times also supports this picture (Mundell et al. 2013).

2.2 Origin of the Prompt Emission

Despite 50 yr of observations, there remain central questions as to the origin of the prompt emission. The most popular model to describe the evolution of the GRB from its formation to late times (both GRB and afterglow) has been the fireball model, first developed by Martin Rees and Peter Mészáros in the early 1990s. Although there are many complexities and additions that have been added to this model over the years, its basic concept remains the same. The premise of this model is that the timescales for the variability seen in GRBs require the deposition of a significant amount of energy in a small volume (this is acutely true in the case where GRBs are cosmological). The fireball is created from the deposition of the heat from some catastrophic event within this small volume. Once released, the significant thermal pressure creates an expansion of the fireball that bores through any material surrounding the region (for example, the remaining envelope of a massive star) before it escapes this material as a relativistic jet with a bulk Lorentz factor of 100 or more. The acceleration of this material requires that the region has little matter around or the energy would become too entrained with that matter, so-called baryon loading, to accelerate it to relativistic velocities.

Although there is little matter entrained within the jet, the high velocities mean there is significant kinetic energy, so much of the thermal energy initially deposited at the base of the jet has been converted. It is likely that the initially released material that may have to escape through some more heavily baryon-loaded material will move at speeds lower than that released later in the process. This difference in velocities of shells of material ejected at different times creates interactions—shocks. Since all of these velocities are very close to the speed of light, these shocks do not occur close to the energy source but instead at much larger radii. This distance can, in principle, be estimated as the time between the ejection of two shells with different Lorentz factors. If these are emitted with a time gap of δt and with Lorentz factors Γ_1 and Γ_2 , then the time in which they will interact is given by $\delta t \Gamma_1 \Gamma_2$, and the distance from the source (given that the velocity is approximately the speed of light) is given by Rees & Meszaros (1994),

$$R_{\text{prompt}} = c \delta t \Gamma_1 \Gamma_2. \quad (2.3)$$

For a time difference of 100 ms and $\Gamma_1 \approx \Gamma_2 \approx 100$, this corresponds to 3×10^{11} m (~ 2 au or $400 R_\odot$). This radius is well beyond the radius of the majority of massive stars, certainly the hydrogen-poor Wolf–Rayet stars that are favored as the progenitors of long GRBs.

Importantly, of course, for shocks to develop, there is a difference between Γ_1 and Γ_2 . As noted in Rees & Meszaros (1994), when the two shocks combine and the momentum between them is shared, their final Lorentz factor $\Gamma_F = \sqrt{\Gamma_1\Gamma_2}$. In this situation, some fraction of the initial kinetic energy is dissipated in the shock, and this is given by

$$\epsilon = \frac{\Gamma_1 + \Gamma_2 - 2\sqrt{\Gamma_1\Gamma_2}}{\Gamma_1 + \Gamma_2}. \quad (2.4)$$

Hence, shocks between shells of different velocities can yield conversion efficiencies in the tens of percent when $\Gamma_1/\Gamma_2 > 2$ (Rees & Meszaros 1994).

In this model, the jet continues through the circumstellar medium but, in doing so, sweeps up the material forming a further external shock between the head of the jet and the surrounding matter. This interaction also creates shocks in which electron acceleration is possible, and, in the fireball model, it is the source of the afterglow emission (see Chapter 3 for more information).

In this standard variant of the fireball model, the emission should be nonthermal in nature and originate from shocks. The presence of thermal components in some GRBs clearly demonstrates that this is not the origin of all the observed prompt emission. Instead, a promising location for the thermal prompt emission arises in the outer photosphere of the fireball. The photospheric radius is defined as Mészáros & Rees (2000); Zhang et al. (2014)

$$R_{\text{ph}} = \frac{L\sigma_T}{4\pi m_p c^3 \Gamma^3} \approx 3.7 \times 10^9 \text{ m} \quad (2.5)$$

and is well within the radius at which shocks are expected to become important.

The very high energy emission also does not apparently originate in the shocks that are created in the fireball model but requires some alternative source that enables relatively delayed emission. A promising, but by no means unique, solution to the origin of the very high energy emission is that of inverse Compton scattering (Beloborodov et al. 2014), most likely off some external medium. In this scenario, the prompt emission (generated from shocks or any other process) creates emission in the keV to MeV range. This is emitted behind the head of the shock wave (the forward shock) but eventually catches it (since the radiation is moving at c and the shock is at a high Lorentz factor but below c). It can then potentially scatter either in the heated plasma behind the forward shock or in the wind of the progenitor star (Beloborodov et al. 2014). Since this material is further out from the engine than the source of the prompt photons and is moving more slowly, Compton scattering within it will more naturally lead to emission that is delayed relative to the bulk of the prompt burst.

2.3 Summary

Given that almost all GRBs detected to date have been found by their prompt emission, and that light curves and spectra, therefore, exist for several thousand bursts, it is perhaps surprising that so many uncertainties remain about the origin of the gamma rays themselves. The details of the physical processes that sculpt the

emission remain uncertain. We do not know the detailed roles of thermal and nonthermal processes and why these appear to differ from burst to burst. There are still questions relating to the origin of the highest-energy emission, and there is not a detailed model available to describe why some bursts appear extremely smooth and others highly variable.

Although much effort in the field has moved away from understanding the details of the prompt emission and now focuses on other areas of GRB physics, there remain central questions to answer, with implications for the progenitors of GRBs and the physics of relativistic outflows. Indeed, while many relations have now been found that may make GRBs valuable as distance indicators (see Chapter 7), the lack of understanding of the physics that drives these relations may yet inhibit progress in that arena.

References

- Ackermann, M., Asano, K., Atwood, W. B., et al. 2010, *ApJ*, 716, 1178
 Band, D., Matteson, J., Ford, L., et al. 1993, *ApJ*, 413, 281
 Beloborodov, A. M., Hascoët, R., & Vurm, I. 2014, *ApJ*, 788, 36
 Boér, M., Gendre, B., & Stratta, G. 2015, *ApJ*, 800, 16
 Briggs, M. S., Band, D. L., Kippen, R. M., et al. 1999, *ApJ*, 524, 82
 Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2013, *ApJ*, 764, 179
 Chattopadhyay, T., Vadawale, S. V., Aarthy, E., et al. 2017, arXiv:1707.06595
 Coburn, W., & Boggs, S. E. 2003, *Natur*, 423, 415
 Gehrels, N., Norris, J. P., Barthelmy, S. D., et al. 2006, *Natur*, 444, 1044
 Gendre, B., Stratta, G., Atteia, J. L., et al. 2013, *ApJ*, 766, 30
 Götz, D., Laurent, P., Lebrun, F., Daigne, F., & Bošnjak, Ž. 2009, *ApJL*, 695, L208
 Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, *ApJL*, 413, L101
 Lei, F., Dean, A. J., & Hills, G. L. 1997, *SSRv*, 82, 309
 Levan, A. J. 2015, *JHEAp*, 7, 44
 Levan, A. J., Jakobsson, P., Hurkett, C., et al. 2007, *MNRAS*, 378, 1439
 Levan, A. J., Tanvir, N. R., Starling, R. L. C., et al. 2014, *ApJ*, 781, 13
 Mazets, E. P., & Golenetskii, S. V. 1981, *Ap&SS*, 75, 47
 McGlynn, S., Foley, S., McBreen, B., et al. 2009, *A&A*, 499, 465
 Mészáros, P., & Rees, M. J. 2000, *ApJ*, 530, 292
 Mukherjee, S., Feigelson, E. D., Jogesh Babu, G., et al. 1998, *ApJ*, 508, 314
 Mundell, C. G., Kopač, D., Arnold, D. M., et al. 2013, *Natur*, 504, 119
 Rees, M. J., & Meszaros, P. 1994, *ApJL*, 430, L93
 Rutledge, R. E., & Fox, D. B. 2004, *MNRAS*, 350, 1288
 Ryde, F. 2005, *ApJL*, 625, L95
 Ryde, F., Axelsson, M., Zhang, B. B., et al. 2010, *ApJL*, 709, L172
 Zhang, B., Zhang, B.-B., Virgili, F. J., et al. 2009, *ApJ*, 703, 1696
 Zhang, B.-B., Zhang, B., Murase, K., Connaughton, V., & Briggs, M. S. 2014, *ApJ*, 787, 66

Gamma-Ray Bursts

Andrew Levan

Chapter 3

Afterglow Emission

The GRBs are defined by their prompt emission, but much of what is known of them arises from studies of their afterglows. Afterglows are necessary to pinpoint GRBs on the sky and measure their redshifts and hence define their energetics, identify their host galaxies, and observe rising supernovae or kilonovae that pinpoint the progenitors of long and short GRBs, respectively. Without the afterglow revolution, begun in 1997 with the discovery of the afterglow of GRB 970228, the origins of GRBs would still be shrouded in mystery.

The search for and discovery of GRB afterglows was not an accident. The models that begin to explain the prompt emission, described in Chapter 2, almost required them. In particular, the spatial distribution of bursts on the sky, available in the early 1990s, ruled out Galactic models for all but very special scenarios and strongly favored cosmological bursts. For these extragalactic models, the GRB properties could only be explained by the creation of a relativistic outflow, and popular models at the time created the GRBs from shocks between material emitted at different Lorentz factors. However, these outflows were not expanding into a perfect vacuum but into some circumstellar medium around the progenitor star. The blast wave itself must contain very little mass to achieve relativistic velocities but has significant energy; thus, as it slows down by plowing into the surrounding medium, it will collect many times its own mass in a shock front. This is often called an external shock, to differentiate it from the internal shocks in the same model that create the prompt emission. The dissipation of energy within this shock will then naturally lead to longer-lived emission at lower photon energies. This is the basic principle of a GRB afterglow and was the signature searched for through the early 1990s.

3.1 The First Afterglow Searches

The appeal of locating GRB afterglows mainly arose because the detection of a burst outside of the gamma-ray window would provide a route to measuring its position on the sky. Gamma rays cannot be focused, even by the grazing incidence optics that

are used for X-ray observations, so it is difficult to accurately localize them. Indeed, in the early 1990s, the coded mask detectors used by, e.g., *Swift* that provide arcminute-level positions from gamma rays alone were not available, and the most accurate positions known for the majority of GRBs were error circles of several degrees in radius. The most precise positions arose from triangulation using a set of detectors spaced across the solar systems and modest additions to other spacecraft—the aptly named InterPlanetary Network (IPN). Since the distance (\equiv time) baseline for these detectors could be very large (e.g., the Earth–Mars distance), the arrival times of gamma-ray photons at different detectors enabled moderately accurate positions to be identified. At times, these were only hundreds of square arcminutes in size, although because the time delay often localized the burst very well in one plane but very poorly in another, these were often narrow strips of the sky that remained difficult to observe. Furthermore, the acquisition and analysis of observations taken as part of the IPN was often slow, and positions were not available for days or sometimes longer after the detection of a given burst.

The difficulty in searching for afterglows in these regions was then compounded by further technological factors. The X-ray satellites were of limited sensitivity compared to those available today and often took a long time to repoint. On the ground, the charge-coupled device (CCD) revolution was well underway, but typically, the resulting chips were relatively small and the field of view commensurately lower than previous wide-field photographic surveys. The location of afterglow emission was therefore strongly motivated, but the difficulties in undertaking such a search should not be underestimated. Nevertheless, various searches were undertaken during the 1990s.

These challenges and the clear scientific returns from precisely located GRBs were major scientific drivers in the development of new satellites that would be able to provide much more precise gamma-ray positions. For example, this was one of the main aims of the *High Energy Transient Explorer (HETE)* satellite that launched in 1996 November. Unfortunately, due to the failure of explosive bolts, the satellite failed to adequately release from a second satellite that was part of the same payload (the Argentinian multipurpose astrophysics satellite *SAC-B*), and the mission was lost. A replacement was funded and launched in 2000, making significant contributions.

However, the major breakthrough arose through observations with the Italian-Dutch satellite *BeppoSAX*, launched in 1996 April. Its wide-field cameras (WFCs) were of a coded mask design, operating in the hard X-ray regime (2–30 keV) and with a field of view of $40 \times 40 \text{ deg}^2$. Such a design uses a mask of tiles above the detector plane, so a source of gamma rays at some position relative to the plane illuminates a unique set of detectors. It is then possible to infer the position based on the set of detectors that “see” a given source. In addition to the WFC, *BeppoSAX* was also equipped with a GRB monitor that could ascertain if any source seen in the WFC was a true GRB. Hence, when the *BeppoSAX* WFC detected a GRB, it was a high-confidence event and could be localized to an error region with a radius of only a few arcminutes (Costa et al. 1997), comfortably within the range of narrow-field X-ray and optical observers.

The crucial breakthrough came on 1997 February 28, when GRB 970228 was detected by the WFC and gamma-ray burst monitor as a classic long-duration GRB with a duration of around a minute. The position was sufficiently accurate (initially 10' but narrowed down to 3' on a timescale of a few hours) that it was possible to observe the location with the narrow-field X-ray telescope in observations taken ~ 8 hr later. These observations, described in full in Costa et al. (1997), revealed a much longer-lived source, which was shown to fade on timescales of several days, appearing a factor of almost 20 fainter in observations taken on 1997 March 3. This was the first X-ray afterglow to a GRB; see Figure 3.1.

A few hours after the first X-ray observations were undertaken, but before they were reported, a series of independent observations of the location of GRB 970228 were performed from the William Herschel Telescope on La Palma (van Paradijs et al. 1997). Unlike X-ray observations, where sources of comparable brightness to GRB afterglows after a few hours are rare, the optical sky is extremely crowded, even within an error box of a few arcminutes. Searches for afterglows therefore cannot identify them with high confidence unless very small (arcsecond-level) positions are known, and sometimes not even in this case. Instead, optical afterglows normally require confirmation that can arise either from the identification of a counterpart that was not present on some pre-explosion image or via the confirmation of variability on images taken some time apart (often consecutive nights). The former approach can be powerful in leading to the rapid identification of the counterpart, important in further follow-up, but it is contingent on the availability of suitable images. While the Digitized Sky Survey, a process to digitize photographic plates, covers the whole sky, it is an old survey, with a shallow depth (typically perhaps equivalent to an R -band magnitude of ~ 20 but variable across its coverage). More recently, far deeper and higher-fidelity images have become available over much (but not all) of the sky, thanks to new generations of wide-field imaging surveys, such as the Sloan Digital Sky Survey, PanSTARRS, and others, which reach magnitudes of $r > 23$. For GRB 970228, a comparison image was not obtained for several nights, with images finally taken by the Nordic Optical Telescope on 1997 March 4. These latter observations showed one source that had vanished at a position coincident with that derived from the narrow-field X-ray observations from *BeppoSAX*. This was the optical afterglow of GRB 970228, the first to be uncovered. Subsequently, deep observations with the New Technology Telescope (NTT) in La Silla, Chile, did unveil an apparently extended source at the afterglow location, the first detection of a host galaxy, although its nature at the time was not entirely uncontroversial (Sahu et al. 1997).

The realization that GRB afterglows could be uncovered spurred many on to attempt localization for future bursts identified by *BeppoSAX* and to undertake deep searches of positions provided by other missions. It motivated the reapproval of *HETE* so that it could contribute to the work, while early work toward the *Swift* concept that would define the field in the 2000s was also begun at this stage. Throughout the late 1990s and early 2000s, afterglows were discovered at a rate of approximately one per month from a range of sources including *BeppoSAX*, *HETE-2*, the IPN, and the *Rossi X-ray Timing Explorer* (*RXTE*). Each event required the

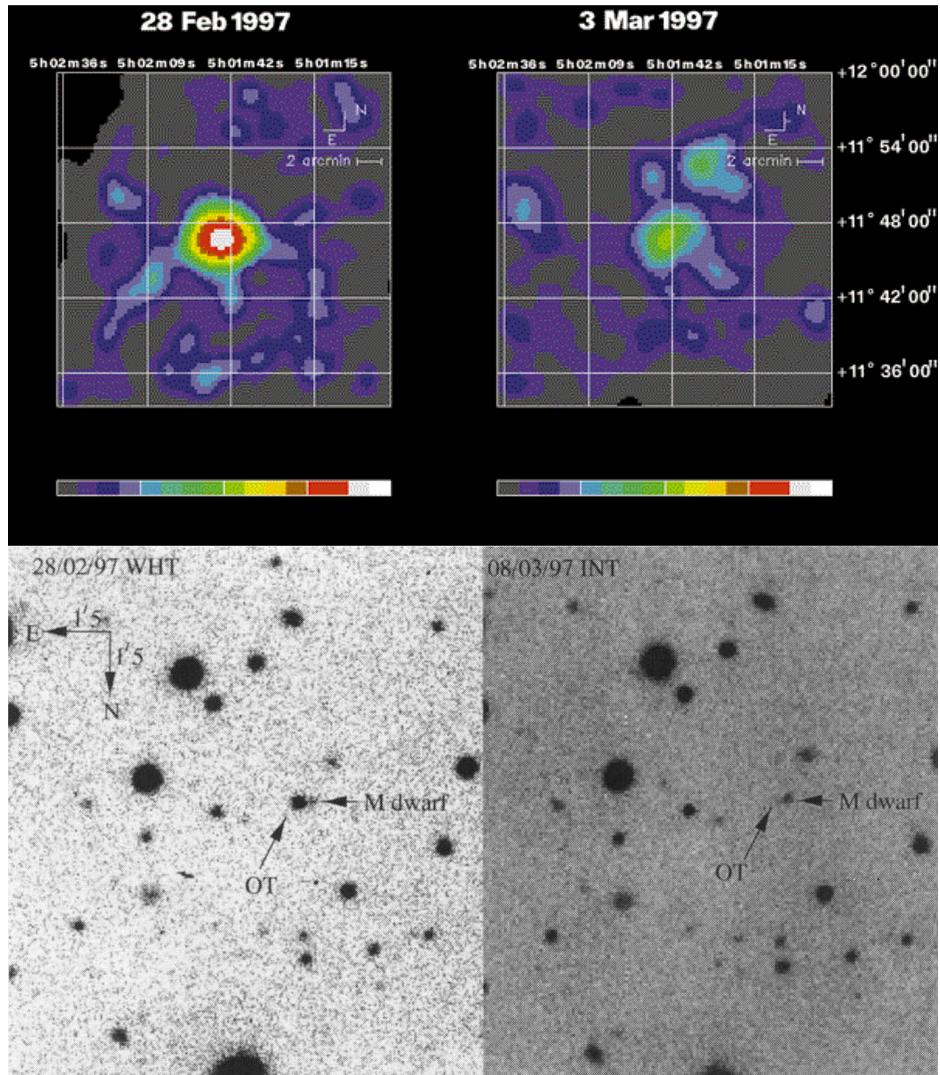


Figure 3.1. Images of the discovery of the X-ray (top; Costa et al. 1997. Reprinted by permission from Springer Nature) and optical (bottom; van Paradijs et al. 1997. Reprinted by permission from Springer Nature) afterglow of GRB 970228, the first GRB for which an afterglow was discovered. Observations on timescales of ~ 1 day and 1 week revealed a fading X-ray and optical source in the region of the sky from which the GRB was identified. It is striking to note that the timescale for the searches in this case was hours to days, whereas GRB searches today are often undertaken in seconds.

significant investment of resources to identify the afterglow and perform further follow-up observations. In part, this effort was related to inexperience on the part of both observers and observatories. Strategies were not well defined, and it was difficult to estimate the necessary depth or cadence of observations or pick the optimal filter. Equally, many observatories did not have procedures in place for dealing with GRB observations on the necessary timescales. As time went by, especially as the *Swift* era began and GRB detection moved from one per month to one every few days, this situation changed, and the majority of major ground- and space-based observatories are now extremely efficient at GRB observing. For example, the Rapid Response Mode on the Very Large Telescope in Chile has resulted in essentially autonomous observations of GRB afterglows less than 10 minutes after a trigger, and smaller robotic telescopes, such as the 2 m Liverpool Telescope, or even smaller facilities, such as ROTSE, have managed the feat even faster, at times when the GRB itself is still ongoing (Racusin et al. 2008).

Several hundred afterglows have now been observed across the electromagnetic spectrum (see Figure 3.2). While it is fair to say that there is strong diversity among

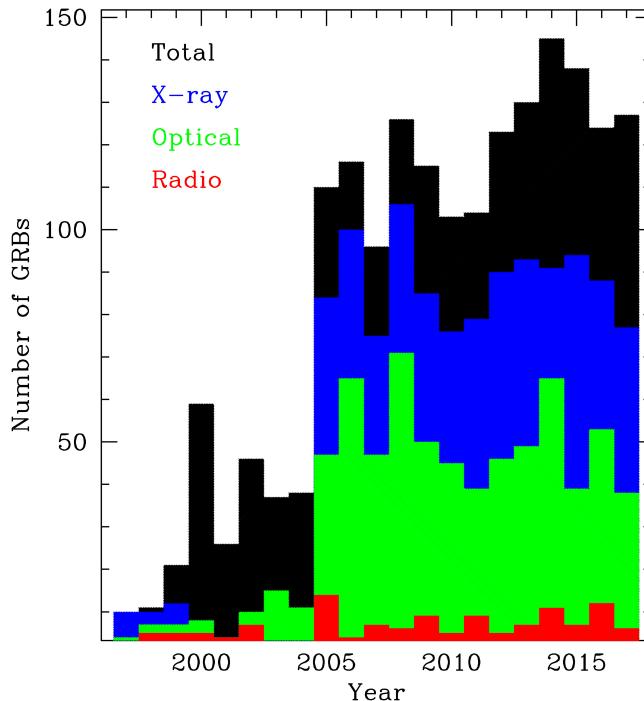


Figure 3.2. Number of GRBs (black) and X-ray (blue), optical (green), and radio (red) afterglows discovered as a function of time since the first afterglows were discovered. Statistics taken from <http://www.mpe.mpg.de/~jcg/grbgen.html>. Note that there is no attempt to remove bursts or interpret them here. For example, many of the bursts without afterglows (especially X-ray afterglows) can be ascribed to a failure to search. The dramatic increase in rate after the launch of *Swift* at the end of 2004 is clearly visible. The recovery fraction is also striking to note, with X-ray afterglows being the most common. Indeed, when a prompt *Swift* slew is possible they are near ubiquitous, followed by optical and, subsequently, radio.

the afterglow population, probably reflecting a diversity in energy, luminosity, shock efficiency, baryon loading, progenitor properties, circumstellar medium, and more, the afterglows of GRBs do appear more similar than the bursts themselves, and it is possible to identify common features within afterglows that lead to some canonical expectations.

3.2 X-ray Afterglows

X-ray afterglows are now almost ubiquitous to GRBs. Although *Swift* does not slew to every GRB it detects because of observational constraints (in particular Earth limb, Moon angle, and Sun angle), in cases where it does, it almost always locates an X-ray afterglow. Less than a handful of the \sim 1000 long GRBs detected by *Swift* and with prompt slews have no X-ray afterglow, and even among the fainter short GRBs, a detection in X-rays is generally expected, even if only a few X-ray photons are ultimately recorded.

3.2.1 The Canonical X-ray Afterglow

X-ray afterglows with *Swift* have now been observed over seven orders of magnitude of brightness and six orders of magnitude in time. While there is significant diversity from burst to burst, the insight from these observations has enabled the description of the light curve of a typical “burst.” A good fraction of all *Swift* GRBs appear to follow the behavior described by this, although the lack of early (or late) observations may mean that some elements are missing. A typical X-ray afterglow is shown in Figure 3.3. Briefly, it can be broken down into several different phases (Nousek et al. 2006; Zhang et al. 2006).

1. *Prompt emission.* For some bursts, XRT observations begin while the BAT is still detecting gamma-ray photons. In this case, the X-ray observations can simply be the soft end of the prompt emission. This is typically extremely bright and can be highly variable (both temporally and spectrally).
2. *Rapid decay.* At the end of the prompt emission, the visible light appears to undergo a rapid decay, often to t^{-5} or steeper. This is most naturally interpreted as arising from high-latitude emission—photons generated at angles away from our direct line of sight that have longer travel times to reach us.
3. *Plateau.* Following the rapid decay, the X-rays often appear to plateau, appearing as flat or with a slow decay ($<t^{-0.5}$). This may relate to ongoing central engine activity.
4. *Flares.* During this plateau phase (or potentially earlier or later), it is possible for bursts to exhibit small or large X-ray flares. In extremum, the energy of these flares may be comparable to the energy in the GRB itself.
5. *Intermediate decay.* At the end of the plateau, the X-ray afterglow will typically begin to fall as t^{-1} . This may occur on timescales of tens of seconds and up to several days in extreme cases. This phase is the one most commonly observed for bursts located prior to the launch of *Swift*.

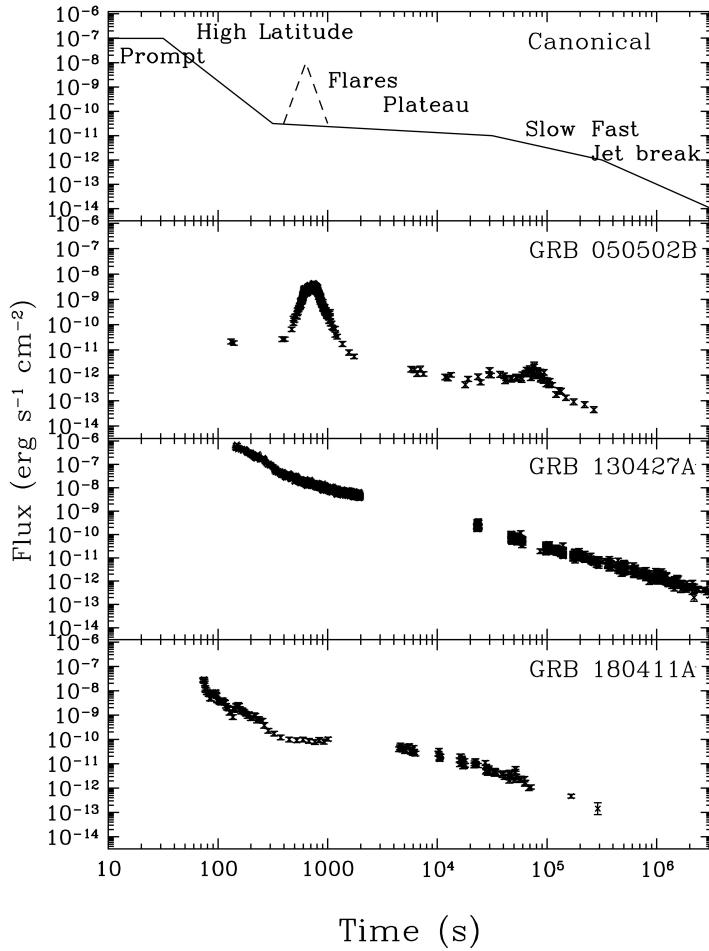


Figure 3.3. Canonical light curve of an X-ray afterglow (top) with typical features marked (Nousek et al. 2006; Zhang et al. 2006). Also shown are several GRBs that exhibit some of these features. In most cases, not all of the features are seen, often due to observational effects (e.g., orbit gaps, later starting observations). In other cases, it is clear that not all features are seen in all light curves. For example, some bursts do not show flares or plateaus.

6. *Rapid decay.* After this slower decay, burst light curves will eventually undergo a temporal break in which the decay steepens. Beyond this point, the flux of the counterpart typically decays as t^{-2} . This temporal break is often referred to as the jet break.

3.3 Optical Afterglows

Optical afterglows were discovered nearly simultaneously with X-ray afterglows. Since the spatial resolution of optical imaging is typically far higher than that for X-ray imaging, optical counterparts in general provide the most accurate positions for GRBs, especially when the goal is to measure this location relative to the host

galaxy or star-forming regions within it. Optical counterparts are also the route through which GRB distances can be obtained, since they illuminate their line of sight from the burst to us, and the imprint of lines from the interstellar medium upon the afterglow enables a direct measurement of the redshift of the absorbing material. Normally, the highest-redshift absorption system (there are frequently more than one, since the line of sight may pass through the extended halos of other galaxies) is allocated as the GRB host redshift, and in general this is likely correct, although formally it is only a lower limit on the burst redshift, as it could arise from a more distant galaxy with little absorption. In a small number of cases, however, variable lines can be seen. These are fine-structure lines such as Fe II (Vreeswijk et al. 2007; D’Elia et al. 2009) and have short lifetimes. They exist because they are “pumped” by UV photons produced in the GRB itself and thus can only be excited if they lie at moderate (kiloparsec) distances from the burst itself. Hence, fine-structure redshifts are direct measurements of the distance to GRBs via their afterglows.

Beyond being a tool for precision positions and redshifts, GRB optical afterglows also provide information relating to the GRB blast wave itself. The afterglows are typically well described by (sometimes broken) power laws in both frequency and time, so that

$$F_\nu(\nu, t) \propto \nu^\beta t^\alpha, \quad (3.1)$$

where $\beta = \Gamma - 1$ and is related to the measured spectral slope often reported in energy from X-ray observations.¹ It is common for $\alpha = -1$ at timescales of hours to days, with a steepening to $\alpha \sim -2$ at later times. Often $\beta \sim -1$, perhaps with a modest degree of curvature introduced by host galaxy extinction. In general, β does not vary with time but is fixed throughout all observations (although a handful of exceptions to this rule do exist).

3.3.1 Dark GRBs

It is striking that while X-ray searches almost always find GRB afterglows when they are performed promptly, the same is not true in the optical and IR. Early in the afterglow era, it was noted that extrapolation of X-rays to the optical (assuming some extrapolation of the X-ray spectral slope) led to the expectation that some optical afterglows should have been detected but were not (Groot et al. 1998). A generic term used for bursts identified at X-ray or radio wavelengths but not in the optical was coined: dark GRBs.

At first, this definition was vague. It was often used simply to describe an optical nondetection but did not use any criteria to determine if the optical observations that failed to detect the afterglow should have done so. For example, many observations were taken very late after the burst, from small telescopes, in poor conditions, or for targets with significant foreground extinction (or sometimes several of the above). In

¹ Here the use of Γ and β refers to the spectral shape of the afterglow, so that $F_\nu \propto \nu^\beta$ and $F_E \propto E^\Gamma$. This notation is used because it is common for the field. It should not be confused with the use of Γ as the bulk Lorentz factor in later sections.

some of these cases, a detection of the optical afterglow may have been almost more surprising than its nondetection. For this reason, several authors sought to quantify dark bursts more meaningfully. For example, Rol et al. (2005) extrapolated X-ray afterglows, accounting for the uncertainties in spectral slope and the possibility of a spectral break known as the cooling break lying between the X-ray and optical regimes (such that $\beta_{OX} = \beta_X - 0.5$; see Section 3.5.2). Perhaps surprisingly, this work found that only a handful of previously identified dark GRBs were actually astrophysically constraining. A simpler “snapshot” identification was proposed by Jakobsson et al. (2004), who simply suggested that $\beta_{OX} < 0.5$. In practice, it is this latter definition that has generally been adopted in the identification of dark GRBs.

The origin of darkness is interesting because of the potential constraints it puts on afterglow models, the nature of the progenitors, the environments of the bursts, and their role as cosmological probes. Optical emission may be absent for a wide range of physical reasons in addition to the observational selection effects outlined above. For example, in most scenarios, the optical luminosity should scale as the particle density, $n^{1/2}$, so bursts with a low circumburst density could drive weak afterglows. At high redshift, the optical light could be redshifted such that observer optical observations probe below the Ly α break where the universe is opaque. If the burst occurred along a very dusty sight line, then it may emit plenty of optical light, but this light may be extinguished within the host galaxy and not reach observers.

In practice, it is likely that all of these effects impact GRBs to some extent. However, it has also become clear that the vast majority of dark GRBs are dark because of absorption in their host galaxies, i.e., because they lie along dusty lines of sight. Evidence for this arises from the high hydrogen column densities measured in their X-ray afterglows (Fynbo et al. 2009; Covino et al. 2013); the presence of “intermediate,” very red, but optically visible bursts (Levan et al. 2006); and host galaxies that appear to be massive and very dusty (Levan et al. 2006; Svensson et al. 2012; Perley et al. 2013). Hence, dark bursts are extremely valuable as cosmological probes but in general are not dark because of intrinsic differences in the properties of their afterglows.

3.3.2 Polarization

The GRB afterglows, like the prompt emission, can also exhibit significant polarization. A purely isotropic outflow would not be expected to yield polarization, so the degree of polarization and its evolution with time provide a direct route to measure the structure of jets within GRBs. Early observations were typically limited to snapshot measurements of polarization, often at times of a few hours to days after the burst. These observations typically yielded low measurements of the polarization, although one measurement of $\sim 10\%$ was recorded (Bersier et al. 2003). With the very bright afterglow of GRB 030329, it was finally possible to obtain high-fidelity, time-resolved polarimetry of the afterglow, although the very complex features resulted in a low ($P < 2\%$) but highly variable degree (and direction) of polarization (Greiner et al. 2003).

With the advent of *Swift*, the possibility of obtaining very early measurements of the polarization also became possible, especially with the presence of polarimeters on telescopes that could respond very rapidly to alerts (e.g., the Liverpool Telescope on La Palma or the Rapid Response Mode implemented at the VLT). The early phases of the GRB may well exhibit strong polarization because of an ongoing contribution from the prompt emission (which apparently can be highly polarized) and because the reverse shock (see Section 3.5.3) may also introduce a unique and powerful polarization singe. The earliest measurements of polarization arise from the Liverpool Telescope (Mundell et al. 2007) and, in the case of GRB 120308A, show a decrease from almost 30% to \sim 15% over the course of a few hundred s (Mundell et al. 2013).

The majority of measurements of GRB polarization have been in the form of linear polarization. However, there is also one detection of circular polarization (Wiersema et al. 2014), which is not naturally expected and remains to be fully explained.

3.4 Radio/Submillimeter Afterglows

Longer-wavelength counterparts to GRBs are rather rarer than those seen in the X-ray or the optical/IR. In part, this is because such observations (at least those reaching the requisite depth) are expensive, so campaigns are not frequently undertaken. Despite this, a good number of radio counterparts have been found, beginning with GRB 970508 (Frail et al. 1997), the burst that also yielded the first GRB redshift.

Unlike the X-ray and optical afterglows, which are typically at their brightest at the time of the first observations, radio afterglows often rise with time, reaching their peak brightness days to weeks after the burst itself. This ameliorates some of the urgency that is necessary for observations at other wavelengths. Radio afterglows are also visible for much longer. Indeed, the detections of some bursts years after the bursts themselves were initially thought to arise from the detection of radio emission from the host galaxy but have been subsequently shown to be the result of long-lived afterglow emission. These radio observations may then be particularly valuable as a means of measuring the energy budget of GRBs, since on timescales of several years, the geometry should have transitioned from highly beamed to effectively spherical, such that the radio afterglows provide a jet-independent route of measuring GRB total energies.

3.5 Emission Processes

3.5.1 Deceleration and External Shocks

As soon as the shock is emitted from the source (in practice, once it escapes its parent star/merger), it will begin to interact with whatever external medium surrounds the progenitor. In order to accelerate the outflow to relativistic velocities, there is very little matter entrained within the jet itself (see Chapter 2), but the extreme Lorentz factors still imply very high kinetic energy. Unlike supernova remnants, which undergo a change in emission from ejecta to ISM domination once they have swept

up the mass of the ejecta from the ISM, GRBs sweep up many times the enclosed mass with relatively little impact on their dynamics. Instead, the GRB forward shock becomes important at the deceleration radius, where the initial Lorentz factor has dropped to half its initial value. If the initial Lorentz factor is given as $\Gamma_0 = E_0/M_{\text{ej}}c^2$, for an explosion with energy E_0 and ejecta mass M_{ej} , then the mass swept up when $\Gamma = \Gamma_0/2$ is given by

$$M_0 = \frac{4\pi}{3} r^3 m_p n \left(\frac{\Gamma_0}{2} \right)^2. \quad (3.2)$$

For the assumption of a constant density medium consisting only of hydrogen (with n particles per unit volume), the radius, r , is often referred to as the deceleration radius r_{dec} and can be simplified to

$$r_{\text{dec}} = 8 \times 10^{16} \left(\frac{E_{53}}{n} \right)^{1/3} \Gamma_0^{-2/3} \text{cm}, \quad (3.3)$$

where E_{53} is the energy release in units of 10^{53} erg (Meszaros & Rees 1993). It should be noted that this distance is extremely large compared to the size of the GRB central engine or even the massive star progenitor. Indeed, while GRBs are extremely compact, their afterglows are generated at far larger radii but appear at the same time as the burst (or, more accurately, shortly thereafter) because they are generated by material traveling toward the observer at nearly the speed of light.

3.5.2 Synchrotron Emission

The dominant emission process that shapes GRB afterglows is the emission of synchrotron radiation from the shock front where the ultrarelativistic outflow impacts the surrounding circumstellar medium about the progenitor. At this shock front, relativistic electrons are accelerated and gyrate about the magnetic field, producing power-law (in frequency and time) afterglow emission that is observed. The details of the observed afterglow then depend on both macrophysical (e.g., burst energy, geometry) and microphysical (e.g., the fraction of the energy in electrons or the magnetic field) properties of the outflow. Since these are potentially complex and variable from burst to burst, the precise properties of the afterglow will then depend upon them. In principle, therefore, a well-mapped afterglow can (nearly) completely define the conditions of the explosion.

Many papers and reviews consider the nature of afterglow emission, and it is only briefly outlined here in its simple form. Variations from the simple models are common and result in varying degrees of additional complexity. Below, the spectral properties are presented as described by Sari et al. (1998); see also Kumar & Zhang (2015).

When the shock wave propagates through the medium, electrons are swept up and shocked, resulting in a distribution of Lorentz factors among the electrons, $N(\gamma_e) \propto \gamma_e^{-p}$. Where $p > 2$ to avoid infinite energy when integrating the electron energy, and is typically taken to be in the region $2 < p < 2.5$. As these electrons

gyrate around magnetic field lines, they emit synchrotron radiation, and this produces emission with a range of photon energies that spans the electromagnetic spectrum. This spectrum is a set of connected power laws with a set of break frequencies corresponding to different physical processes within the emitting region.

At low frequencies, the source essentially behaves as a blackbody because emitted photons are reabsorbed. This occurs at frequencies lower than the so-called self-absorption frequency, ν_a . Below this frequency, the dependence on the specific flux per unit frequency (F_ν) with frequency is given as $F_\nu \propto \nu^2$. At higher frequencies, the emission scales as $F_\nu \propto \nu^{1/3}$ up to the peak (or maximum) frequency, ν_m . Here ν_m is determined by the characteristic frequency for synchrotron radiation, which is given by Kumar & Zhang (2015),

$$\nu_{\text{syn}} \approx \nu_m = \frac{\omega_{\text{syn}}}{2\pi} \sim \frac{qB\gamma_e^2}{2\pi m_e c}, \quad (3.4)$$

where γ_e refers to the electron Lorentz factor, m_e to the mass of the electron, and B to the field strength. At frequencies higher than the peak, the flux is dominated by the sum of the emission from electrons with Lorentz factors larger than γ_e at the peak, and the resulting spectrum is $F_\nu \propto \nu^{-(p-1)/2}$ and thus typically $\nu^{-1/2}$.

The final break frequency is related to the cooling of the electrons as they gyrate around the field, the so-called cooling frequency. The electrons are shock-accelerated but cool as they radiate. Initially, only those with high Lorentz factors cool, but as time continues, progressively more electrons have cooled (so the cooling frequency is not fixed in time but moves to lower frequencies). At frequencies higher than the cooling frequency, the emission scales as $F_\nu \propto \nu^{-p/2}$. The expected resulting broadband (radio to X-ray) spectral energy distributions are shown in Figure 3.4.

The task in modeling the GRB afterglow is therefore to pin down the location of the spectral breaks, their evolution with time, and the spectral slopes within each regime. For most GRB afterglows observed at reasonable times (hours to days after the burst), it appears that $\nu_c < \nu_m < \nu_a$. Here ν_c is often lying between the X-ray and optical regimes, ν_m in the submillimeter to radio, and ν_a always in the radio frequency range.

Once this is done, these simplest GRB afterglow models can be fit to extract the physical parameters of the explosion itself, most notably, the electron injection index, p ; the particle number density in the surrounding medium (n , or more formally the radial density profile, which may be constant or vary as r^{-2} for a wind-like medium); the energy, E ; and the electric and magnetic energy fractions of the electrons, ϵ_e and ϵ_B .

3.5.3 The Reverse Shock

In addition to the outgoing forward shock that plows into the interstellar medium, there is also a so-called reverse shock, which propagates backward through the ejecta. In many ways, “reverse” shock can be misleading; this shock is still moving outward from the rest frame of the star at relativistic velocity but is plowing backward through the ejecta in their frame and slowing the expansion. This reverse

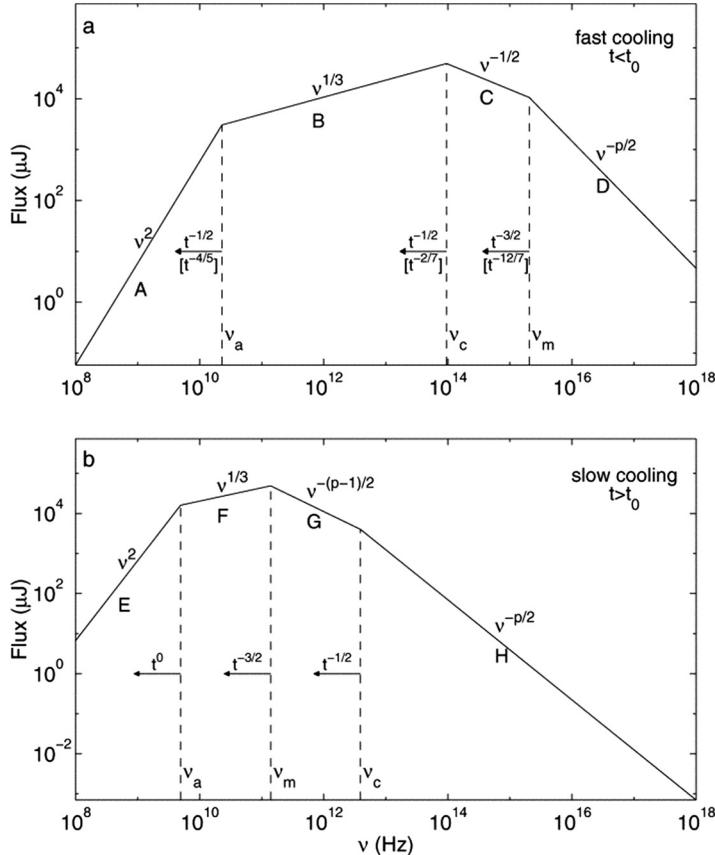


Figure 3.4. Predicted synchrotron spectral energy distributions for GRB afterglows from Sari et al. (1998). The two regimes of fast (early time) and slow (later time) cooling are shown. The majority of afterglows at the times typically observed fall in the slow-cooling regime, and the cooling break lies between the optical and the X-ray. Numerous observations support this broad picture for afterglows. © 1998. The American Astronomical Society. All rights reserved. Printed in USA.

shock can be dynamically important, as it can carry comparable energy to the forward shock (Sari & Piran 1999). However, it should emit predominantly at frequencies well below the gamma-ray regime because of far lower temperatures. Indeed, the difference in the peak frequency of the synchrotron emission from the reverse and forward shock is a factor of Γ^2 (Laskar et al. 2013). Therefore, a broad expectation for reverse-shock emission is the presence of bright optical/IR and radio flares in early GRB afterglows. These flares are not important at later times in the emission because, while the forward shock continues to propagate into the surrounding medium essentially indefinitely, the reverse shock ceases to be important once it has traversed the ejecta.

The identification of signatures of the reverse shock is challenging. They occur early in the afterglow when the forward shock is strong and for long bursts while the prompt emission may continue. A clear identification thus requires the accurate

decoupling of these differing emission processes, and hence comprehensive observations across the electromagnetic spectrum from early to late times. It is likely that the optical flash observed in a handful of GRBs, such as GRB 990123 (Akerlof et al. 1999), is due to the passage of the reverse shock, and a robust detection has been found in early radio-to-X-ray observations of GRB 130427A (Laskar et al. 2013; see Figure 3.5).

3.6 Evidence for Relativistic Beaming

The above discussion is valid for isotropic emission. However, for the early phases of the GRB afterglow, it still provides a good description even if the GRB is highly collimated, since the individual emitting regions of the outflow are not in causal contact at large angles and thus behave as though they are expanding isotropically (see below). If the emission from the GRB itself is confined to a relativistic jet with some half opening angle, θ_j , then the true burst energy is not that observed by a GRB detector but is modified by the fraction of the sky illuminated by the jet. This is only strictly true for a so-called top-hat jet, where the energy per solid angle is the same for any observer, but is commonly assumed. In this case, the correction from the measured energy of the burst E_{iso} to its true energy E_γ is given by Frail et al. (2001),

$$E_\gamma = E_{\text{iso}}(1 - \cos \theta_j) \approx E_{\text{iso}} \frac{\theta_j^2}{2}, \quad (3.5)$$

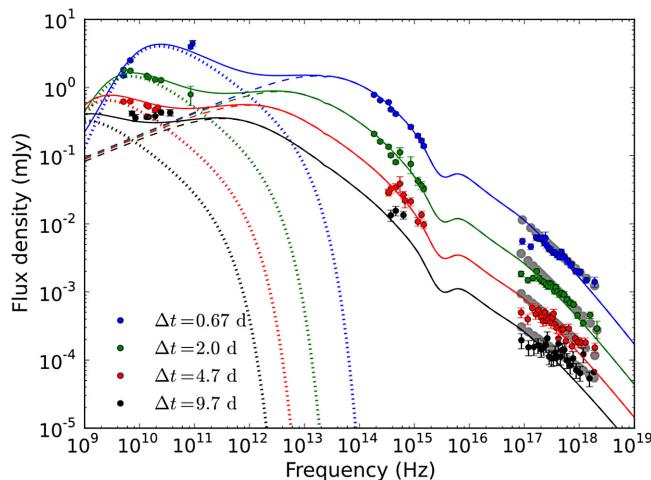


Figure 3.5. Spectral energy distribution of the afterglow of the very bright GRB 130427A from Laskar et al. (2013). The bluer light (optical and X-ray) appears to follow a typical synchrotron forward-shock expectation (e.g., Sari et al. 1998; note that the apparent features in the X-ray and optical spectra are due to the presence of dust within the host galaxy). However, this model provides a poor fit to early radio observations, which lie systematically above expectations. Instead, these data are explained by the presence of a much cooler (lower-frequency-emitting) reverse shock. In this case, the high-quality multiwavelength coverage enabled the burst afterglow to be decoupled into the forward (solid lines) and reverse (dashed lines) shock components. © 2013. The American Astronomical Society. All rights reserved.

where the geometric term is often expressed as the beaming fraction $f_b = (1 - \cos \theta_j)$. The impact of this beaming is clearly significant, since $\sim 5^\circ$ jets imply a modification of the total energy budget by a factor of 250 combined with the same factor increase in the true astrophysical rate. Hence, there is significant importance in identifying the true beaming angles of GRBs, and the best route to achieving this comes from late-time multiwavelength monitoring of their afterglow to search for so-called jet breaks.

At early times, individual elements within the relativistic outflow are not in causal contact, since they experience time dilation. However, the jet is slowing as it plows into the medium, and the evolution of the Lorentz factor is given by $\Gamma = \Gamma_0(t/t_0)^{-3/2}$, where Γ_0 is the initial Lorentz factor at t_0 . As time passes and the jet slows down, it eventually reaches the point where $\Gamma = 1/\theta_j$. At this point, the two sides of the jet are in causal contact. Beyond this crucial point, the afterglow is expected to fade more rapidly, yielding a steepening of the slope and a so-called jet break.

It is therefore broadly agreed that the jet break provides a route to measuring GRB beaming and occurs when $\Gamma = 1/\theta_j$. However, the physical processes that shape this change are somewhat less clear. In principle, beyond this point, the Lorentz factor declines exponentially with radius, yielding a much lower energy input (Rhoads 1999). The jet should feel its pressure, to the point that it becomes causally connected and should expand sideways (Granot 2007). Finally, there are geometric effects; as the jet edges become visible and the observer integrates over a region of emission to obtain a brightness, there is no additional emission from outside the jet cone, so the source appears to fade more rapidly (van Eerten et al. 2012).

Several well-studied GRBs show strong evidence for jet breaks (see Figure 3.6), which manifest themselves as achromatic breaks in the power-law decay of the afterglow (i.e., they occur at X-ray, optical, and radio wavelengths simultaneously). Typically, the decay rate of the afterglow changes from $\sim t^{-1}$ to $\sim t^{-2}$ over the course of the jet break, such that following GRBs beyond the break becomes increasingly difficult due to the rapid decay.

3.6.1 Jet Structure

The evidence for an expanding relativistic jet in GRBs is extremely strong, but the detailed physical parameters of the relativistic jet remain unclear. The simplest model is of a so-called top-hat jet, where the energy (or Lorentz factor) is the same across the jet and drops to zero at sharp edges. To first order, for these simple models, any observer who was within the cone of the GRB jet would then see the “same” GRB. This model is clearly not entirely realistic, and the edges of any associated jet are likely to slow due to interaction with the surrounding medium (for example, as they plow through a star); however, it may be reasonable to first order. Alternatively, the jet may show significant structure. For example, it may exhibit a much higher Lorentz factor when viewed directly down the jet axis, but this may decrease rapidly as the viewing angle increases. The functional form of the Lorentz factor with viewing angle is uncertain and could have many forms. It is common in

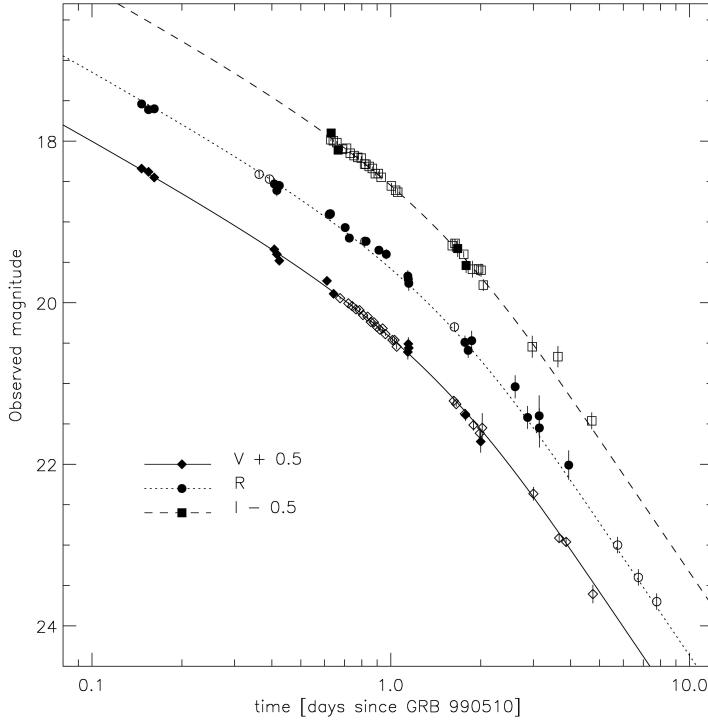


Figure 3.6. Early example of a jet break in a GRB afterglow, GRB 990510 (Harrison et al. 1999). The clear, achromatic steepening of the afterglow is interpreted as the jet break at the point that $\Gamma = 1/\theta$, and provides a direct route for measuring GRB collimation. © 1999. The American Astronomical Society. All rights reserved. Printed in USA.

structured jet models to assume that the jet has a Gaussian structure or falls as the square of the viewing angle (Nakar et al. 2004).

In practice, the jet structure dramatically impacts the properties of the bursts. If jets are strongly variable, then the range of apparent GRB energetics observed may be related to the range of viewing angles of what may be very similar jets (burst to burst). Alternatively, if the jets are uniform, then the range of GRB properties must somehow reflect the properties of the jet itself (intrinsic energy, opening angle, etc).

The structure also impacts the appearance of the jet to viewers at later times. In particular, in the post-jet-break regime, the jet becomes visible to viewers at progressively larger angles, including those who were not originally within the beam. This means that, in principle, it is possible to observe a GRB afterglow (from either a structured or uniform jet) when a GRB was not seen. These are so-called orphan afterglows (GRB afterglows without an associated GRB; Granot et al. 2002; Levinson et al. 2002). There have been significant searches for such events, and indeed some candidates have been found (Cenko et al. 2013, 2015), although it remains challenging to determine if, for a given event, the GRB was genuinely absent or simply missed. Interestingly, the recent short GRB associated with the gravitational wave-detected binary neutron star merger GW 170817 does appear to

have an off-axis GRB afterglow (Troja et al. 2017; Margutti et al. 2017; Lyman et al. 2018), although in this case, the detection of a weak GRB suggests that it cannot be considered a true “orphan” afterglow.

3.6.2 Simulations

Simulations of GRBs are challenging for a number of reasons. First, the range of scales involved challenges many computational methods. The central engine itself is perhaps only 10 km across, but the afterglow is produced at perhaps 10^{16} cm or more. Similarly, the central engine properties vary on timescales of milliseconds, but radio afterglows must be tracked for years. These issues of dynamic range essentially preclude simple simulations in which grid sizes or timesteps are in some way fixed and require more sophisticated approaches in which they can be varied. Furthermore, there is a requirement for simulations to account for the effects of both general and special relativity to accurately constrain both the properties of the central engine and the resulting outflow. Indeed, the complexity of the simulations means that it is still not possible to model in detail the evolution of the star, its collapse (or merger for a short GRB), and the resulting outflow, and indeed, the tools to do each stage are not necessarily compatible. Instead, different approaches can be taken depending on the desired study.

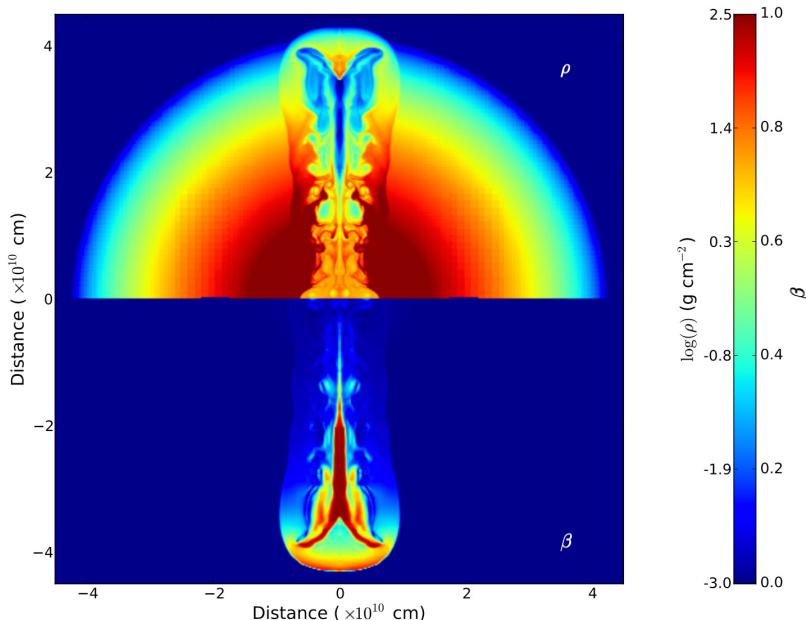


Figure 3.7. Simulations of a GRB jet (from Lazzati et al. 2012). The figure shows a jet 8 s after the engine became active (and was active for 7.5 s) and just as the jet clears the star (the so-called breakout). The upper panel shows the density of the material, while the lower panel shows the velocity (here $\beta = v/c$). The creation of a low-density, high-velocity jet is clear, although there is also significant structure at the edges in the form of turbulent eddies. © 2012. The American Astronomical Society. All rights reserved.

For the study of afterglows, the primary concern is to investigate the nature of the outflow from its source to large radii and then to integrate the properties into some form of observable. In principle, this can be done analytically (see, e.g., Granot 2007), but such analytical approaches are not without their problems and fail to capture issues, such as turbulence, that only become visible when multidimensional calculations are undertaken. In the past few years, relativistic hydrodynamical calculations have reached the level of sophistication where it is possible to model the expansion of a jet (given as some source of energy at the core) through the progenitor star and directly predict the observable quantities. Such simulations are clearly a vital complement to both observations of GRB afterglows and the analytical solutions described above. Indeed, such simulations can potentially reveal details of important physical processes, such as the jet–star interaction, that may impact the subsequent evolution of the afterglow (through, for example, imprinting structure on the jet). Snapshots of one simulation are shown in Figure 3.7, where the jetlike structure on breakout from the progenitor star can clearly be seen.

References

- Akerlof, C., Balsano, R., Barthelmy, S., et al. 1999, *Natur*, **398**, 400
 Bersier, D., McLeod, B., Garnavich, P. M., et al. 2003, *ApJL*, **583**, L63
 Cenko, S. B., Kulkarni, S. R., Horesh, A., et al. 2013, *ApJ*, **769**, 130
 Cenko, S. B., Urban, A. L., Perley, D. A., et al. 2015, *ApJL*, **803**, L24
 Costa, E., Frontera, F., Heise, J., et al. 1997, *Natur*, **387**, 783
 Covino, S., Melandri, A., Salvaterra, R., et al. 2013, *MNRAS*, **432**, 1231
 D'Elia, V., Fiore, F., Perna, R., et al. 2009, *ApJ*, **694**, 332
 Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M., & Taylor, G. B. 1997, *Natur*, **389**, 261
 Frail, D. A., Kulkarni, S. R., Sari, R., et al. 2001, *ApJL*, **562**, L55
 Fynbo, J. P U., Jakobsson, P., Prochaska, J. X., et al. 2009, *ApJS*, **185**, 526
 Granot, J. 2007, RMxAA, **27**, 140
 Granot, J., Panaiteescu, A., Kumar, P., & Woosley, S. E. 2002, *ApJL*, **570**, L61
 Greiner, J., Klose, S., Reinsch, K., et al. 2003, *Natur*, **426**, 157
 Groot, P. J., Galama, T. J., van Paradijs, J., et al. 1998, *ApJL*, **493**, L27
 Harrison, F. A., Bloom, J. S., Frail, D. A., et al. 1999, *ApJL*, **523**, L121
 Jakobsson, P., Hjorth, J., Fynbo, J. P U., et al. 2004, *ApJL*, **617**, L21
 Kumar, P., & Zhang, B. 2015, *PhR*, **561**, 1
 Laskar, T., Berger, E., Zauderer, B. A., et al. 2013, *ApJ*, **776**, 119
 Lazzati, D., Morsony, B. J., Blackwell, C. H., & Begelman, M. C. 2012, *ApJ*, **750**, 68
 Levan, A. J., Davies, M. B., & King, A. R. 2006, *MNRAS*, **372**, 1351
 Levinson, A., Ofek, E. O., Waxman, E., & Gal-Yam, A. 2002, *ApJ*, **576**, 923
 Lyman, J. D., Lamb, G. P., Levan, A. J., et al. 2018, *NatAs*, **2**, 751
 Margutti, R., Berger, E., Fong, W., et al. 2017, *ApJL*, **848**, L20
 Meszaros, P., & Rees, M. J. 1993, *ApJ*, **405**, 278
 Mundell, C. G., Kopač, D., Arnold, D. M., et al. 2013, *Natur*, **504**, 119
 Mundell, C. G., Steele, I. A., Smith, R. J., et al. 2007, *Sci*, **315**, 1822
 Nakar, E., Granot, J., & Guetta, D. 2004, *ApJL*, **606**, L37
 Nousek, J. A., Kouveliotou, C., Grupe, D., et al. 2006, *ApJ*, **642**, 389

- Perley, D. A., Levan, A. J., Tanvir, N. R., et al. 2013, *ApJ*, **778**, 128
Racusin, J. L., Karpov, S. V., Sokolowski, M., et al. 2008, *Natur*, **455**, 183
Rhoads, J. E. 1999, *ApJ*, **525**, 737
Rol, E., Wijers, R. A. M. J., Kouveliotou, C., Kaper, L., & Kaneko, Y. 2005, *ApJ*, **624**, 868
Sahu, K. C., Livio, M., Petro, L., et al. 1997, *Natur*, **387**, 476
Sari, R., & Piran, T. 1999, *ApJ*, **520**, 641
Sari, R., Piran, T., & Narayan, R. 1998, *ApJL*, **497**, L17
Svensson, K. M., Levan, A. J., Tanvir, N. R., et al. 2012, *MNRAS*, **421**, 25
Troja, E., Piro, L., van Eerten, H., et al. 2017, *Natur*, **551**, 71
van Eerten, H., van der Horst, A., & MacFadyen, A. 2012, *ApJ*, **749**, 44
van Paradijs, J., Groot, P. J., Galama, T., et al. 1997, *Natur*, **386**, 686
Vreeswijk, P. M., Ledoux, C., Smette, A., et al. 2007, *A&A*, **468**, 83
Wiersema, K., Covino, S., Toma, K., et al. 2014, *Natur*, **509**, 201
Zhang, B., Fan, Y. Z., Dyks, J., et al. 2006, *ApJ*, **642**, 354

Gamma-Ray Bursts

Andrew Levan

Chapter 4

Central Engines

4.1 The Requirement of a Central Engine

Early observations of gamma-ray bursts made it apparent that the GRB emission must be arising from a compact region. Simple time-delay arguments suggested that the emitting region could only be of order $c\Delta T$ across, and for millisecond variability, this corresponds to sizes of hundreds of kilometers at most. For cosmological explosions with high peak fluxes (for example, 10^{-4} erg cm $^{-2}$ s $^{-1}$), the photon density close to the source must be large, and pair production should be significant. In principle, this means that the source region should be optically thick and create little emission beyond the rest-mass energy of an electron, $E = m_e c^2 = 0.511$ MeV. This problem was recognized early in the GRB field and often described by a so-called compactness parameter (Cavollo & Rees 1978),

$$C = \frac{L\sigma_T}{m_p c^3 R} \approx \frac{F_\gamma d^2 \sigma_T}{m_p c^4 \Delta T} \sim \frac{10^{12} F_{\gamma-4} d_{\text{Gpc}}^2}{\Delta t(\text{ms})}, \quad (4.1)$$

where σ_T is the Thompson cross detection, d is the luminosity distance (in Gpc when expressed as d_{Gpc}^2), and F_γ is the gamma-ray flux (in units of 10^{-4} erg cm $^{-2}$ s $^{-1}$ in cases where $F_{\gamma-4}$ is given). If $C \gg 1$, then it suggests that the source is optically thick and should not be capable of producing a GRB. Indeed, in this case, the spectrum of the observed emission would be purely thermal, in contrast to the apparent nonthermal emission that is observed.

At first sight, this would suggest that GRBs cannot be compact, highly energetic, cosmological sources, since for any modest peak flux, $C \gg 1$. In practice, this concern is alleviated because GRBs are relativistically expanding. In this case, the observed energy of the photons is boosted by the Lorentz factor Γ , so that their rest-frame energies are often well below the pair production threshold for an assumed Lorentz factor in the region of $\Gamma \sim 100$. This, in turn, means that while the source region where the energy is produced is very compact, the distance at which the gamma rays are actually created is much larger, perhaps by factors of 10^{12} or more.

These observations directly imply that GRBs arise from physically small volumes in which there is a large-scale deposition of energy that is sufficient to accelerate ambient material to very close to the speed of light. This is the role of the central engine.

Indeed, it is now apparent that long-duration GRBs arise from the collapse of massive stars and that short GRBs (or at least some of them) are created from the merger of two compact objects (either neutron star–neutron star or neutron star–black hole). In these systems, the natural central engine is the newly formed compact object in the core. This may be a black hole or a highly magnetized neutron star, and it may derive the energy to power the explosion in a variety of ways. Some of these are outlined below.

4.2 Black Hole Central Engines

The first and still most commonly considered central engines in GRBs are those that use a newly formed black hole as the engine. This newly formed black hole can be created directly in the core collapse of a massive star or via the merger of two neutron stars if the mass of the resulting remnant exceeds the maximum mass of the neutron star (see Figure 4.1). This maximum mass is poorly understood, since it depends on the details of the neutron star equation of state, but it is likely to be close to twice the canonical $1.4 M_{\odot}$ that many neutron stars are found to have. Such stellar mass black holes are found with typical masses of several to $10 M_{\odot}$ and, within the Galaxy, are normally found as companions in compact binary systems where their dynamical masses can be directly measured.

Once the black hole has formed, the most promising route to extracting energy is from material surrounding the black hole that forms an accretion structure. In a massive star, this likely requires the star to be sufficiently rapidly rotating that some of the material within the stellar core has too much angular momentum to be captured within the black hole on formation, while in the neutron star merger scenario, it is likely to arise because of tidally distorted material that is expelled from each neutron star in the final few orbits prior to the merger. Formally, one can determine that there must be some material that has sufficient specific angular momentum (angular momentum per unit mass) to form a centrifugally supported disk at the innermost stable orbit of the black hole. Formally,

$$j > \frac{\sqrt{6} GM}{c}, \quad (4.2)$$

which is typically of order $10^{16} \text{ cm}^2 \text{ s}^{-1}$ (Levan et al. 2016). Such angular momentum is relatively straightforward in the case of a merging neutron star, but the majority of massive stars observed in the Milky Way will fall short by an order of magnitude or more (see Chapter 5).

Once this structure is in place, it is necessary to extract the energy to create the GRB. The exact process by which this is done remains uncertain, but a strongly discussed possibility is that the GRB itself is created by neutrino–antineutrino annihilation from neutrinos created from the hot disk. This is possible since, despite

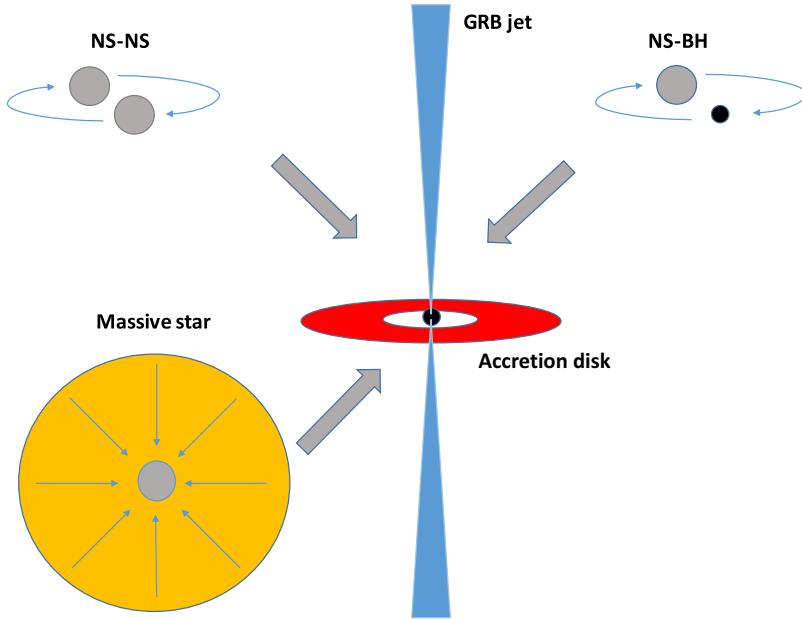


Figure 4.1. Formation of a black hole accretion disk system in the black hole central engine model for GRBs. The accretion disk is formed because at the time of collapse, some fraction of the material has sufficient angular momentum that it cannot accrete directly but instead forms a disk at the innermost stable orbit. This can potentially be formed from disparate physical systems, such as merging neutron stars (which have significant orbital angular momentum) or a rapidly rotating massive star. This provides a natural explanation for some of the similarities between GRBs created via very different progenitor channels. For a more complete census of routes to creating black hole accretion disk systems, see Fryer et al. (1999).

the extremely low interaction rate of neutrinos to terrestrial detectors, the cross section for weak interactions scales as the square of the temperature; thus, in the high-density and high-temperature regions, the path length is sufficiently short to enable energy deposition via annihilation. Since this annihilation occurs above and below the disk, it produces a natural preference for the expulsion of material along the polar axis. Indeed, there is an angular dependence on the annihilation rate (Nagataki 2018) that, given the accretion disk geometry, produces the highest interaction rate above and below the black hole (roughly speaking, the maximum is when the cosine of the angle between the two neutrinos is $\cos \theta_{\nu\nu} = 0$). Hence, this annihilation leads to deposition of energy in the right place to accelerate the jets.

Alternatively, the GRB jet may be generated by electromagnetic effects, in particular the tapping of the rotation of the black hole via electromagnetic (EM) fields in the surrounding accretion disk via the so-called Blandford–Znajek effect (Blandford & Znajek 1977). This is a commonly considered route to the creation of astrophysical jets at a range of scales, in particular in quasars, and could also operate in GRBs. However, the formation of astrophysical jets remains an open question in GRBs and in the field more generally and is one that is difficult to simulate numerically or observe directly.

Once the jet is formed, its lifetime and luminosity are expected to be linked directly to the accretion rate onto the newly formed black hole. It is generally thought that the accretion rate will be highest (in the region of $M_{\odot} \text{ s}^{-1}$) at early times and will then rapidly decrease as the disk is depleted. Provided the jet remains powered at the source for the time it takes to break out from the dense region (the star in the case of a long GRB), a GRB will be observed. In practice, it is unlikely that accretion ceases rapidly in black holes within massive stars, although its rate may drop significantly. Eventually, any accretion is likely to be fueled by fallback accretion of material initially ejected but with insufficient velocity to escape the potential well. This fallback rate depends on the details of the explosion, although for generic expectations, it should scale as $t^{-5/3}$. Fallback accretion is an attractive explanation for the origin of long-lived plateaus in the afterglows of many GRBs. Furthermore, it is possible that material within the ejecta will clump, rather than raining down in a consistent matter. The degree of fragmentation is unclear and may occur either in the disk via so-called Toomre instabilities (Perna et al. 2006) or in the supernova itself (King et al. 2005). The accretion of such fragments may cause the bright flares that are seen in some GRBs.

4.3 Magnetar Central Engines

An alternative to black hole accretion models is that the GRB energy is provided by a highly magnetic neutron star: a magnetar, whose dipole spin-down and neutrino emission immediately after formation are sufficient to create the energy to power the GRB. Such models have been discussed since the early 1990s (Usov 1992; Thompson & Duncan 1993) and have come to prominence again in recent years (Metzger et al. 2011, 2015).

Magnetars are well documented within our galaxy, and while no strong definition is agreed upon, a widely accepted interpretation is a surface magnetic field of $B > 10^{14} \text{ G}$. To date, there are approximately 30 such examples known in the Milky Way (Olausen & Kaspi 2014). Their fields are typically inferred from their spin period (P) and its derivative (\dot{P}), $B \propto P\dot{P}$. Similarly, since the spin-down is dominated by the field, the time to spin down from an effectively zero spin period to the one observed provides an estimate of the characteristic age of the magnetar, $\tau = P/2\dot{P}$, which for galactic systems is typically several thousand yr.¹ The rotational energy of a spinning neutron star is

$$E_{\text{rot}} = \frac{1}{2}I\omega^2, \quad (4.3)$$

where ω is the angular frequency and $I = 2/5 mr^2$ is the moment of inertia for a sphere of mass m and radius r . Assuming that magnetars are born with spin periods of perhaps 1 ms (close to the breakup rotation rate for a neutron star), these magnetars have lost approximately 10^{52} erg of energy since their formation. If such

¹This is naturally an approximation, both because the neutron stars may not start with an essentially zero period and because the spin-down rate actually depends on the so-called braking index. While this can be defined for pure magnetic braking, it appears unlikely that pulsars or magnetars always follow this.

energy can be extracted rapidly, then it could readily power many GRBs, with only the most energetic examples potentially challenging this model.

In order to be relevant for GRB formation, the time to spin down to periods of seconds cannot be the thousands of years observed for most magnetars we observe in our Galaxy but must be of order tens to hundreds of seconds. In particular, the rate of energy loss is given by

$$\frac{dE_{\text{rot}}}{dt} = \frac{d}{dt}\left(\frac{1}{2}I\Omega^2\right) = I\Omega\dot{\Omega} = -\frac{4\pi^2 I \dot{P}}{P^3}, \quad (4.4)$$

since $\Omega = 2\pi/P$. We can see from this that maximizing the energy deposition rate requires a high spin-down rate (\dot{P}) and a small period (P). The short period is expected after a supernova and is indeed required to carry the total requisite energy to power a GRB. As the dipole field $B \propto (P\dot{P})^{1/2}$, the requirement of spin-down within a few hundred seconds and an initial period of milliseconds also fixes the possible range of magnetic fields, which are typically in the regime of $B \sim 10^{15}$ G.

Once this magnetar is formed and begins to deposit its energy, it is still necessary to channel this energy into a GRB jet and create the observables that are found in GRBs. As with the black hole model, there are some uncertainties as to exactly how this may be done. However, descriptions of the mechanisms at play do exist.

As shown in Figure 4.2 and described in detail in Metzger et al. (2011), the GRB production can be broken down into several stages. Briefly, as the star collapses and the neutron star is formed, a neutrino-driven wind is pushed by the proto-neutron star and becomes relativistic on a timescale of a few seconds. It may naturally be directional because of preferential neutrino mass loss along open field lines, but it can also then be collimated by the star, depending on its own structure (e.g., if there is a preferential axis through which escape is easier, perhaps because the rapid rotation has left the polar regions more rarefied during the collapse). This jet punches through the remainder of the star in the same way as a jet driven in any other model, and once the jet breaks out of the stellar (or supernova) envelope, there is a clear channel from the magnetar directly through the progenitor through which further jet activity can be observed. Once it has escaped the star, the jet creates the observed GRB, which is active as a GRB for as long as the burst is powered. It has internal variability governed by processes occurring close to the neutron star and may be created by either magnetic dissipation or shocks. As noted by Metzger et al. (2011), shocks are inevitable because the wind from the neutron star starts at subrelativistic velocities and then becomes faster, such that later emission will naturally catch up with the earlier material. At some later point, the engine becomes transparent to neutrinos, and the jet power stops. Once the jet has opened a cavity in the star that is visible to outside observers, these observers should view changes in observable properties that are directly linked to the activity of the central engine, as there is no other material blocking the line of sight or interacting with the radiation. Once the jet ceases to be active, the energy should be dominated by the spin-down of the magnetar. This is valuable because it means that when comparing the early afterglows of GRBs to the magnetar model, it is possible to do so simply by describing the evolution of its dipole field and spin and the resulting radiation.

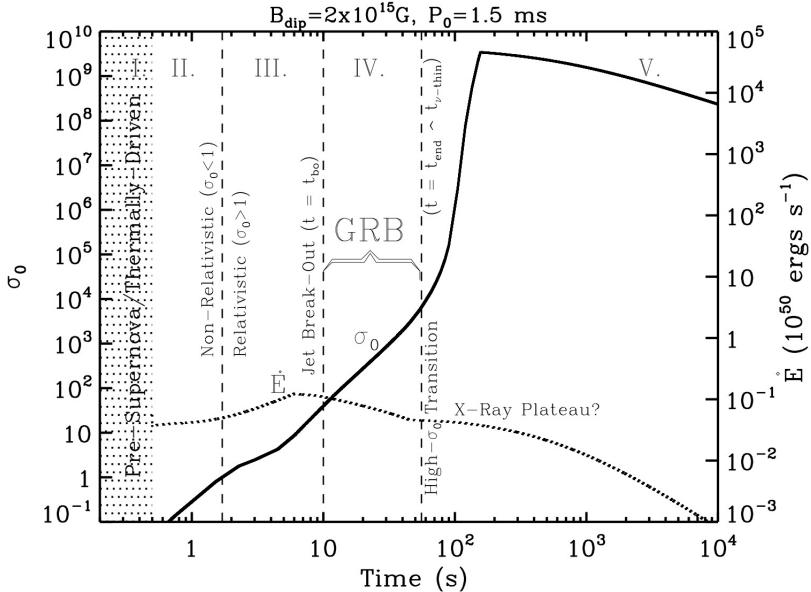


Figure 4.2. Evolution of the energy output (\dot{E}) and the magnetization $\sigma_0 = \phi \Omega^2 / \dot{M} c^3$ (where $\phi = Br^2$ is the magnetic flux) of a protomagnetar with time (from Metzger et al. 2011). Stage I refers to the formation of the magnetar during the supernova, after which a neutrino-driven wind is formed (stage II). This transitions rapidly to a relativistic wind (stage III) and then breaks out of the star, creating the visible GRB (stage IV). Once the system becomes transparent to neutrinos, the jet power stops and the viewer observes the dipole spin-down and a possible X-ray plateau (stage V). Reproduced from Metzger et al. (2011). By permission of Oxford University Press on behalf of the Royal Astronomical Society.

Of particular interest is the presence of plateaus in the X-ray afterglows of many GRBs observed by *Swift*. Many of these bear a similarity to the predictions of the magnetar model at these late stages, where the spin-down is dominant. In particular, since $\dot{E} = -4\pi^2 I \dot{P} / P^3$ and $B \propto (P \dot{P})^{1/2}$, the energy output at this time is given roughly as $\dot{E} \propto B^2 P^{-4}$ (Metzger et al. 2011). This means that the change in the energy output is dominated purely by the spin-down of the magnetar, creating a relatively slow plateau. In most GRBs, this plateau ends at times of a few hundred s. This may be the spin-down time of the magnetar. Beyond this spin-down time, the energy input should vary as t^{-2} or, indeed, could cease if the magnetar was hypermassive and centrifugally supported to start with, since it may collapse to form a black hole at this point. The X-ray flares can also be explained in this model, since these young magnetars may well undergo outbursts following similar physical mechanisms to those found in galactic magnetars. To date, there is no evidence of periodicity in GRB X-ray light curves, which would be a hallmark of this model, but the spin periods are sufficiently short that such searches are not yet sufficiently sensitive to be constraining (Rowlinson et al. 2017).

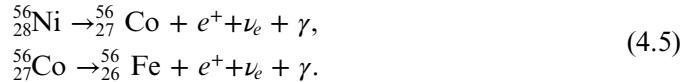
It is worth noting that in this model, there can be no formation of significant accretion disks, or at least no significant accretion while the magnetar is active. In

particular, the addition of mass at early times could readily drive the magnetar beyond the maximum mass of a neutron star and hence result in black hole formation and no magnetar-driven explosion. In principle, it may be possible to avoid such accretion if the magnetic field of the magnetar is sufficiently strong that it acts to drive any (charged) material away. Indeed, accretion in highly magnetic binaries appears to be suppressed by this propeller action, which limits the accretion rate.

4.4 Central Engines in Other Astrophysical Transients

The vast majority of supernovae should create either neutron stars or black holes, and only those from extremely massive and low-metallicity stars that explode as so-called pair-instability supernovae should yield no remnant at all. Normal supernova explosions are much more common than those that create GRBs, so it is apparent that in the majority of cases, the newly formed compact object does not have a significant and immediate impact on the resulting explosion.

Magnetar models have proved very popular and successful in explaining the nature of the so-called superluminous supernovae (SLSNe; Kasen & Bildsten 2010). The SLSNe reach peak magnitudes brighter than $M_V \sim -21$ (Gal-Yam 2012), a factor of several tens to hundreds brighter than typically seen for core-collapse events (that have mean $M_V \sim -17$) and a factor of 10 brighter than Type Ia supernovae (which have $M_V \sim -19$). Such luminous supernovae are difficult to explain via traditional models, which posit that the optical luminosity around peak is driven by the radioactive decay of nickel via the β -decays,



which have half-lives of 6.1 and 77.7 days. The bolometric luminosity of the supernova is then limited by the number of ${}^{56}_{28}\text{Ni}$ atoms created (i.e., the nickel mass). However, in order to explain SLSNe, it is necessary to invoke several solar masses of ${}^{56}_{28}\text{Ni}$, rather than the few tenths at most created in the majority of supernovae. This is only possible in extremely massive stars. This led to suggestions that SLSNe may be the long-sought-after pair-instability supernovae, in which the high temperatures and pressures in the core of extremely massive, low-metallicity stars spontaneously produce electron–positron pairs. The creation of these pairs at the expense of highly energetic, relativistic photons lowers the central pressure in the star, which causes it to contract, enhancing nuclear-burning rates. This can lead to either the direct explosion of the star in a pair-instability supernova or the reestablishment of nuclear burning, in which case the star may collapse in stages via a so-called pulsational pair-instability supernova. Since these pair-instability events occur in stars with initially very high masses and high rates of nuclear burning, it may be possible to create sufficiently bright supernovae to explain the SLSNe, and several notable examples have been suggested to arise from this mechanism (Gal-Yam et al. 2009).

However, the late-time energy injection from radioactivity is strongly driven by the nuclear decays themselves, with a fixed half-life and thus an expected decay. Indeed, the hallmark of a radioactively powered supernova is a late-time light curve

that follows the expected exponential decay rates. This is not true for the majority of SLSNe, so alternative models are needed. Here the magnetar model is particularly appealing (Kasen & Bildsten 2010). In particular, the total energies of SLSNe are well within the energy budget of a magnetar, and certain parameters yield a spin-down time of a few weeks, comparable to the peak of the supernova and the photon diffusion time through the ejecta. Since the energy input from the magnetar follows a straightforward analytical form, it is possible to derive direct expressions for the luminosity and peak time of a magnetar-driven explosion as a function of the magnetar (although it does, of course, also depend on the ejecta mass and energy; Kasen & Bildsten 2010). Fits to these parameters provide a remarkably good explanation of the light curves of many SLSNe (Inserra et al. 2013).

It is interesting to note that there is a modest overlap of magnetar parameters in which it is possible to both power a supernova and create a GRB. In most cases, these two are distinct, because the spin-down time in the GRB case is ~ 100 s, and in the case of SLSNe, it is days to weeks. Rapid spin-down is unlikely to create significant boosts to the luminosity of a supernova because the energy deposited by magnetars at this point is probably either in the GRB jet or the kinetic energy of the supernova (i.e., it may boost the velocity and explain, at least in part, why GRB supernovae are high-velocity events, but it is unlikely to have such an impact on the luminosity). However, for magnetars with rapid spin periods but modest magnetic fields, spin-down times of $\sim 10^4$ s are possible. In this scenario, it may be possible to create a so-called ultralong GRB (GRBs with durations of several thousand s; Levan et al. 2014) and a modestly enhanced supernova luminosity. There is evidence that at least one such event may exist in the form of the ultralong GRB 111209A (Levan et al. 2014; Gendre et al. 2013), which is associated with a luminous supernova and with a peak luminosity lying between that of GRB supernovae and SLSNe (Greiner et al. 2015). In this sense, it appears that there is a broad range of highly energetic astrophysical transients that can be explained by magnetars with a diverse but physically plausible range of properties; see Figure 4.3 and Metzger et al. (2015).

More recently, work has investigated whether the presence of magnetars within more “normal” supernovae could have significant impacts on their appearance (Sukhbold & Thompson 2017). The majority of the 30 or so magnetars within the Milky Way so far detected are young systems, with ages of $< 10^4$ yr in most cases. This implies that magnetars are in fact a common result of supernova explosions. The supernova rate is typically considered to be of order once per century, so in the past 10^4 yr, we expect of order 100 supernovae; of these, perhaps 20%–30% have created magnetars, and this could be higher given that our census of magnetars is far from complete. It is therefore possible that some fraction of magnetars do in fact impact their supernovae in significant ways, and they can produce reasonable matches to the light curves of several types of hydrogen-rich core-collapse events.

One issue with the magnetar model is that the light curves of supernovae are frequently relatively smooth and can be described by models with a modest number of parameters. The magnetar model is therefore often capable of fitting the broad appearance of a light curve, and this does not necessarily imply that one must be present. Stronger constraints on magnetars can be found by higher-quality

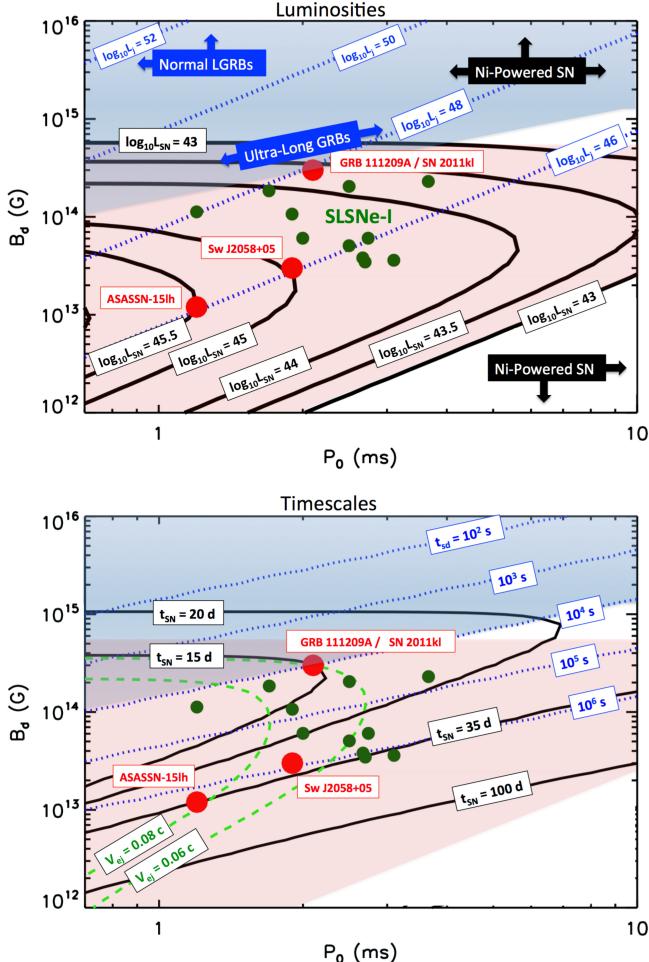


Figure 4.3. Parameter space for the creation of different kinds of magnetar-powered explosions from Metzger et al. (2015). The axes for the two figures are the same, showing the period and magnetic field of the nascent magnetar. Dashed lines in the top panel indicate the resulting luminosity, demonstrating how high-luminosity (but short-lived) events such as GRBs are preferentially created from short-period, high-field magnetars, while ultralong GRBs and SLSNe are created from events with longer periods and lower fields. The bottom panel has dashed lines that indicate the spin-down times for the same magnetars. As can be seen, the very rapidly spinning, high-field magnetars must also spin down rapidly, creating a consistent picture of GRB and SLSNe production. Reproduced from Metzger et al. (2015). By permission of Oxford University Press on behalf of the Royal Astronomical Society.

observations at late times, where the energy injection rates from magnetars show stronger divergence both from radioactive decay and black hole accretion and from directly attempting to explain multiple properties via magnetars. This is most notable in GRBs, where it can be tested whether the magnetar properties are consistent with the properties of both the GRB and the resulting supernova. Indeed,

this is challenging, since the timescales required for powering each are very different, and thus may place relatively strong constraints on the physical properties required for the magnetar itself.

Finally, magnetars have a firm upper limit to the available energy: the rotational energy of a neutron star rotating at breakup. This is not a trivial number to determine, since it depends on both the mass and radius of the neutron star and hence its equation of state (something relatively poorly constrained by observations at this time). However, it seems likely that the total energy emitted by the magnetar can rarely be greater than a few times 10^{52} erg, with some of this lost in the efficiency of the conversion of rotational energy into electromagnetic light. Any system whose energy greatly exceeds this value must (i) be highly beamed, so that its true energy is reduced below this threshold; (ii) have an additional source of energy (e.g., accretion, radioactive power from a supernova); or (iii) not be formed via magnetar.

Central engines, in particular those drawn from black holes, may also be visible in other highly energetic astrophysical transients. Perhaps the most striking example is that of the tidal disruption event (TDE) in which a luminous transient with a duration of several weeks is created by the disruption and subsequent accretion of a star by a supermassive black hole (Rees 1988; or, potentially, a white dwarf by an intermediate-mass black hole (Krolik & Piran 2011)). Many examples of such systems have now been identified by wide-field surveys, and it is apparent that they occur at rates of perhaps once every 10^5 yr for a typical galaxy. Interestingly, it is also apparent that, like GRBs, at least some fraction of these systems also create powerful and long-lived (weeks) relativistic jets (Levan et al. 2011; Bloom et al. 2011; Burrows et al. 2011; Zauderer et al. 2011), such that they can be viewed as significantly scaled-up versions of GRBs (e.g., jets from the accretion of solar masses of material but onto a supermassive rather than stellar mass black hole).

4.5 Summary

It is clear that compact central engines, be they black holes or neutron stars, can have important consequences for the appearance of many classes of astrophysical transients, not least GRBs. Indeed, GRBs are perhaps the best example of an engine-driven explosion, because the properties of the bursts themselves directly imply that only compact objects can enable the necessary conditions for their formation. While the emission from GRBs is generated at large radii (due to the relativistic expansion of GRB jets), the properties of the GRB can link back directly to the properties of the central engine. This is very unusual in most astrophysical systems, where the cores are shrouded by material and cannot be probed directly (with the possible exception of gravitational waves in certain circumstances). Hence, GRBs provide a unique route to understanding engine-driven explosions across cosmic history and their role in shaping the universe we observe today.

References

- Blandford, R. D., & Znajek, R. L. 1977, *MNRAS*, 179, 433
 Bloom, J. S., Giannios, D., Metzger, B. D., et al. 2011, *Sci*, 333, 203

- Burrows, D. N., Kennea, J. A., Ghisellini, G., et al. 2011, *Natur*, **476**, 421
- Cavallo, G., & Rees, M. J. 1978, *MNRAS*, **183**, 359
- Fryer, C. L., Woosley, S. E., & Hartmann, D. H. 1999, *ApJ*, **526**, 152
- Gal-Yam, A. 2012, *Sci*, **337**, 927
- Gal-Yam, A., Mazzali, P., Ofek, E. O., et al. 2009, *Natur*, **462**, 624
- Gendre, B., Stratta, G., Atteia, J. L., et al. 2013, *ApJ*, **766**, 30
- Greiner, J., Mazzali, P. A., Kann, D. A., et al. 2015, *Natur*, **523**, 189
- Inserra, C., Smartt, S. J., Jerkstrand, A., et al. 2013, *ApJ*, **770**, 128
- Kasen, D., & Bildsten, L. 2010, *ApJ*, **717**, 245
- King, A., O'Brien, P. T., Goad, M. R., et al. 2005, *ApJL*, **630**, L113
- Krolik, J. H., & Piran, T. 2011, *ApJ*, **743**, 134
- Levan, A., Crowther, P., de Grijs, R., et al. 2016, *SSRv*, **202**, 33
- Levan, A. J., Tanvir, N. R., Cenko, S. B., et al. 2011, *Sci*, **333**, 199
- Levan, A. J., Tanvir, N. R., Starling, R. L. C., et al. 2014, *ApJ*, **781**, 13
- Metzger, B. D., Giannios, D., Thompson, T. A., Bucciantini, N., & Quataert, E. 2011, *MNRAS*, **413**, 2031
- Metzger, B. D., Margalit, B., Kasen, D., & Quataert, E. 2015, *MNRAS*, **454**, 3311
- Nagataki, S. 2018, *RPPh*, **81**, 026901
- Olausen, S. A., & Kaspi, V. M. 2014, *ApJS*, **212**, 6
- Perna, R., Armitage, P. J., & Zhang, B. 2006, *ApJL*, **636**, L29
- Rees, M. J. 1988, *Natur*, **333**, 523
- Rowlinson, A., Patruno, A., & O'Brien, P. T. 2017, *MNRAS*, **472**, 1152
- Sukhbold, T., & Thompson, T. A. 2017, *MNRAS*, **472**, 224
- Thompson, C., & Duncan, R. C. 1993, *ApJ*, **408**, 194
- Usov, V. V. 1992, *Natur*, **357**, 472
- Zauderer, B. A., Berger, E., Soderberg, A. M., et al. 2011, *Natur*, **476**, 425

Gamma-Ray Bursts

Andrew Levan

Chapter 5

Long-GRB Progenitors

5.1 The GRB–supernova Connection

The strongest evidence for the origin of long GRBs arises from their connection with hydrogen-poor Type Ic supernovae. Type II supernovae are the most common core-collapse event and show clear hydrogen emission features and a variety of spectral and temporal evolution. Type I supernovae do not show any such hydrogen. Type Ib supernovae exhibit helium lines, while Type Ic are devoid of both hydrogen and helium. The GRBs are associated with the Type Ic variety and, furthermore, appear exclusively in Type Ic supernovae with broad lines (so-called SNe Ic-BL), indicative of a high expansion velocity ($\sim 30,000 \text{ km s}^{-1}$; Cano et al. 2017).

Evidence for the origin of GRBs in supernovae rapidly became available following the first detections of the afterglows, most notably via the detection of SN 1998bw associated with GRB 980425, only a year after the first afterglow identifications (Galama et al. 1998). This burst remains an enigma; it is by some margin the closest robustly identified long GRB ever observed ($z = 0.0085$ or ~ 35 Mpc) and was extremely weak (a total isotropic equivalent energy of 10^{48} erg, compared to 10^{54} erg for more typical GRBs). Soon after its detection, it was apparent that the 8' radius error box of GRB 980425 (Pian et al. 2000) contained the local galaxy ESO 184-G82. Observations taken in the following nights found not a fading afterglow at this location but rather a rising supernova. Indeed, there is no evidence of an optical afterglow at all; while X-ray and radio observations do point to a scenario in which some relativistic material was released, this was clearly much weaker than most GRBs (Kouveliotou et al. 2004). Spectroscopy of this supernova showed a spectrum that had marked similarities to those of previously observed hydrogen- and helium-deficient supernovae (Type Ic) but exhibited very broad lines consistent with a much higher kinetic energy and hence expansion velocity, with $\sim 30,000 \text{ km s}^{-1}$ observed during the early phases. The burst appears to originate from within a weak but clear star formation region within the host galaxy. As the supernova evolved, it was also notable that it was unusually luminous, reaching a peak magnitude of $M_B \sim -19.3$,

comparable to a Type Ia supernova, and perhaps an order of magnitude brighter than most core-collapse events. The absence of such supernovae in previous supernova searches suggested a very rare event, unlikely to arise by chance within the relatively small gamma-ray localization, and hence it was generally accepted that GRB 980425 was indeed causally related to SN 1998bw.

Given this discovery, efforts were put into place to search for such signatures in other GRBs. All GRBs observed are more distant than GRB 980425, and the majority are too far for any supernova to be readily observed (the mean redshift is $z > 2$; Jakobsson et al. 2006, 2012). However, SN 1998bw was bright, and so it was possible to observe its signature in GRBs out to $z \sim 1$ and even to obtain spectroscopic observations of reasonable quality for bursts in the $0.1 < z < 0.5$ range. The crucial photometric signature expected in a more distant GRB is that the decline of the afterglow should slow or even reverse as the supernova contribution rises. Simultaneously, we would expect the color of the source to evolve from the blue power-law spectrum of the afterglow to a redder thermal spectrum from the supernova. Claims of such signatures were made in several bursts in the following months to years. These included the likely detection of supernovae in bursts observed prior to GRB 980425/SN 1998bw, such as GRB 980326 (Bloom et al. 1999), although this lacks a clear redshift, and GRB 970228 (Galama et al. 2000). More tailored observations were obtained for newly discovered low- z bursts, although the rate of detection of these was small. However, convincing evidence for the presence of supernovae with bulk properties similar to those of SN 1998bw was found in several bursts, including GRB 011121 (Garnavich et al. 2003).

There were also claimed supernovae based on typically low signal-to-noise spectroscopy in some cases, and if interpreted correctly, these suggested that a range of ejecta velocities was possible, with GRB 021211 (Della Valle et al. 2003) providing a better match to the spectrum of the low-velocity Type Ic SN 1994I. However, such spectral features were not unambiguous because of the difficulty of disentangling the components of the observed light, which arises from the sum of the afterglow $F_{\text{AG}}(\nu, t)$, supernova ($F_{\text{SN}}(\nu, t)$), and host (F_{host} ; Levan et al. 2016), e.g.,

$$F_{\text{obs}}(\nu, t) = F_{\text{AG}}(\nu, t) + F_{\text{SN}}(\nu, t) + F_{\text{host}}(\nu). \quad (5.1)$$

While the host can ultimately be removed, high-quality observations are required to disentangle the afterglow and supernova components, and measurements of the afterglow in bands where the supernova contribution is expected to be minimal are valuable. This might, for example, entail observations in the ultraviolet, where most supernovae are faint due to metal line blanketing, or even in the X-ray. Since it is normally assumed that the afterglow is a power law in both time and frequency, its measurement in different bands (free from supernova contamination) in principle allows its contribution to the supernova to be removed.

Ideally, a burst for supernova studies would look rather like SN 1998bw. It would lie at low redshift, where the supernova could be detected; have a weak GRB component, so that afterglow contamination was not a problem; and, in an ideal world, originate from a faint host galaxy or one that could be resolved such that the underlying stellar light caused minimal confusion.

Needless to say, such bursts are clearly rare, so the majority of events had few of these advantages. Indeed, it remains the case that none are as close as GRB 980425. Further, while analogs to SN 1998bw/GRB 980425 were valuable for studies of the associated supernovae, they would be extreme outliers for GRBs, so there was a significant drive to undertake an in-depth supernova search in a cosmological, high-luminosity GRB. A major step forward was possible with observations of the *HETE-2*-detected burst GRB 030329. This burst was extremely bright, but this was mainly because of a low redshift of $z = 0.17$, with the isotropic energy of GRB 030329 lying at $E_{\text{iso}} \approx 10^{52} \text{ erg s}^{-1}$ (Kouveliotou et al. 2004), consistent with the bulk of the GRB population (although with an energy at the lower end of the distribution). However, the low redshift enabled major spectroscopic campaigns to be launched. Over the course of the following weeks, these revealed that the light evolved from a pure power law into a spectrum with unmistakable similarities to that observed for GRB 980425 (see Figure 5.1). This provided conclusive evidence of the GRB–supernova connection. Interestingly, it also demonstrated the necessity of spectral observations to be confident of the properties of the supernova. The light curve of GRB 030329 is highly variable, and from this alone, it would be extremely challenging to identify or isolate a supernova component (Lipkin et al. 2004).

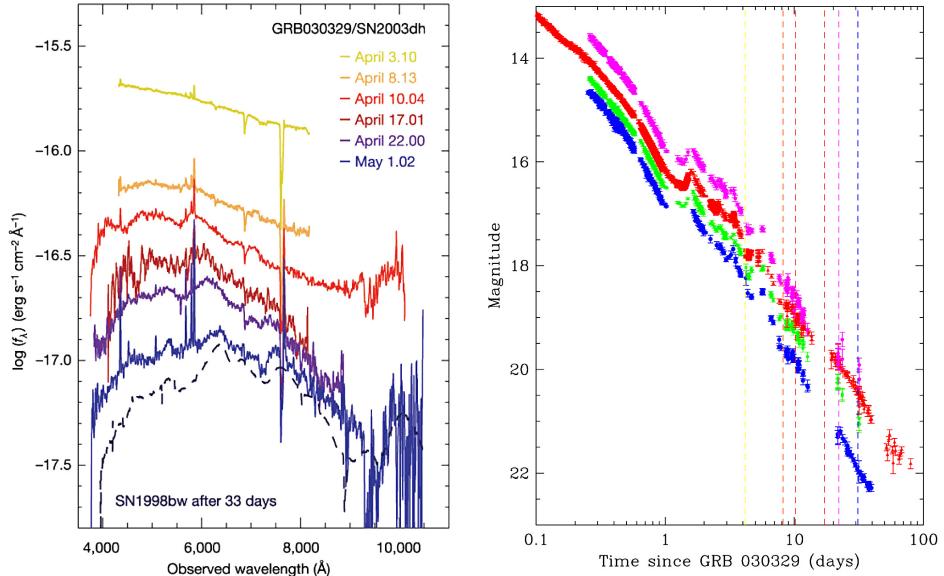


Figure 5.1. Spectra (left; Hjorth et al. 2003, Reprinted by permission from Springer Nature) and light curves (right; adapted from Lipkin et al. 2004) of the optical counterpart of GRB 030329. The spectral sequence clearly shows the evolution from a pure power-law afterglow at early times to a supernova comparable to SN 1998bw at later epochs, sealing the association between GRBs and highly energetic, hydrogen-poor supernovae. However, it is striking for this event that the same signature is not as visible in the light curves. The dashed lines indicate the epochs that spectroscopy was undertaken. It can be seen that there is a reddening of the afterglow, but the complex, non-power-law behavior at early times would complicate the interpretation in the absence of spectral observations.

Indeed, this is a clear example where the simple expectation of an afterglow that is a power law in time and frequency is wrong, and where additional components—for example, clumpiness in the surrounding medium or the refreshing of the forward shock by ongoing engine activity—significantly impact the observed properties.

Interest in the GRB–supernova connection has continued to be a major area of research since these discoveries, and the 15 yr since have provided a much clearer view of the properties of associated GRB supernovae, although bursts at $z < 0.3\text{--}0.5$, where such studies can be undertaken well, remain rare. Nonetheless, there are now several well-studied examples of GRB supernovae, both from local low-luminosity bursts and from the more typical cosmological population. A thorough review of the details of GRB supernovae is given in Cano et al. (2017), and the ensemble properties of the majority of GRB supernovae can be summarized relatively straightforwardly.

Spectrally, GRB supernovae are spectrally similar but not identical (see Figure 5.2). All GRB supernovae are of Type Ic (showing no sign of either hydrogen or helium), and (almost) all exhibit broad lines from the same range of species (when there is sufficient signal-to-noise to observe them); in particular, this includes O I, Ca II, Si II, and Fe II (Cano et al. 2017). The velocity widths of these lines at early times are in the tens of thousands of km s^{-1} and do not resemble the majority of Type Ic supernovae that exhibit much lower ejecta velocities. There are a handful of broad-lined Type Ic supernovae observed without a GRB. In a small number of these cases, there does appear to be a misaligned jet (Soderberg et al. 2010), suggesting that a GRB may have been visible to a suitably aligned observer, but this is clearly not the case in most events, since the jet energies are not comparable.

Photometrically, there is some variation in the light-curve shape and peak magnitude of GRB supernovae. The brightest, SN 2011kl, associated with the ULGRB 111209A (see below and Greiner et al. 2015), peaks at an absolute magnitude of $M_V \sim -20$ and is also spectrally distinct from other GRB supernovae, with a spectrum appearing closer to those seen in superluminous supernovae (Gal-Yam 2012; see also Chapter 4). At fainter peak magnitudes, there is a modest dispersion of perhaps a factor of 2–3 between the brightest and faintest GRB supernovae, with a typical peak magnitude in the region of $M_V \sim -19$. For the handful of bursts followed to suitably late times, the decay appears to follow the expected exponential form for a supernova powered by nickel, with inferred masses of a few tenths of a solar mass of ^{56}Ni produced in each event. It has been suggested that as with Type Ia supernovae, the peak brightness of a GRB supernova correlates with its decay rate, such that brighter events decay more slowly (Cano 2014). This may enable the supernova to be standardized and used for direct measurements of luminosity distance, although such an approach has yet to prove valuable as either a measurement or a confirmation of cosmological parameters.

5.1.1 The Collapsar Model

The most popular route to explaining the connection between long GRBs and Type Ic supernovae is the collapsar model, first developed in the early 1990s (Woosley

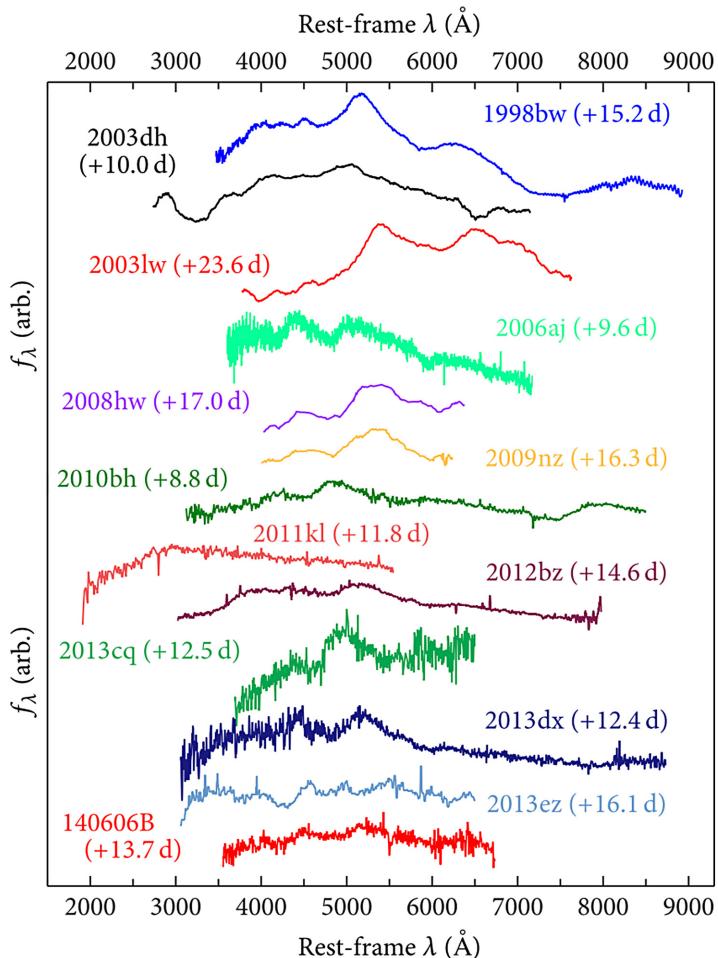


Figure 5.2. Sample of peak time spectra of GRB supernovae (from Cano et al. 2017). The prototype GRB supernova, SN 1998bw/GRB 980425, is shown at the top, along with several other examples of varying spectroscopic quality. With the exception of SN 2011kl, the broad description of the spectral features is similar, although there are more subtle distinctions between the spectra. The features are extremely broad and evolve rapidly (in velocity) with time. Only SN 2011kl (GRB 1111209A) appears markedly different from these expectations, with a very blue spectrum. Notably, while other GRB supernovae (and, indeed, many supernovae) show a strong suppression of light blueward of 3000 Å, SN 2011kl does not, and instead rises to a peak blueward of this value. This is typical of the superluminous supernovae. Reproduced from Advances in Astronomy. CC BY 4.0.

1993). This model envisaged the creation of a black hole with a surrounding accretion disk as the central engine of the GRB (see Chapter 4). This engine provides the necessary energy reserve to power the burst, is likely to create polar jets to explain GRB collimation, and explains the association with supernovae and star-forming regions.

Within the collapsar model, the expected progenitor is a Wolf–Rayet star, which is compact and hydrogen-poor, consistent with the lifetimes of GRBs and the absence

of hydrogen in GRB supernovae spectra. The core evolution of the star proceeds to silicon burning and creates an iron core and then ceases to produce energy because of the unfavorable binding energy for nucleons for fusion beyond iron. At this point, core collapse occurs as in other supernovae, but the mass is sufficiently high that a black hole is formed, either promptly (i.e., on infall) or by fallback. However, because of the rapid rotation of the progenitor, some material has sufficient angular momentum that it cannot accrete directly and forms an accretion disk at the innermost stable orbit of the black hole. This disk then fuels the creation of the GRB (see Chapter 4), and, in the original proposal for the collapsar model, this was done via $\nu\bar{\nu}$ annihilation from the inner regions of the accretion disk.

The expectation of collapsars is that they would arise from stars significantly more massive than those that create typical supernovae due to the necessity of black hole formation. The requirement for rapid rotation, which most massive stars do not have at the end of their lives, would also explain why GRBs are apparently so much rarer than supernovae.

The collapsar model offers a remarkably good description of long GRBs, although it does not specify how the progenitor star forms in detail, only its final conditions. It is also worth noting that the original collapsar model envisaged failed supernovae, in that it was not obvious that while the collapse occurred it was also possible to drive a strong outward shock and supernova explosion, although earlier work had noted that the accretion of significant ($\sim M_\odot$) mass from the inner regions was still consistent with the launching of a supernova (Bodenheimer & Woosley 1983). In any case, the collapsar model, as it now applies to the vast majority of long GRBs, does involve the creation of a bright, hydrogen-poor supernova.

5.1.2 Long GRBs without Supernovae

The emerging picture from long GRBs appeared to be secure in the early 2000s, with supernovae visible in all GRBs where searches were plausible and undertaken. This suggested that there was a common mechanism at play in the creation of all GRBs. However, two GRBs discovered in 2006 cast doubt on this scenario. Both GRB 060505 and GRB 060614 are low-redshift events ($z = 0.089$ and 0.125 , respectively) where an associated supernova should be straightforward to observe. However, in both cases, there is no sign of any supernova emission to limits a factor of up to 100 times fainter than SN 1998bw and a factor of >10 fainter than most core-collapse supernovae (Della Valle et al. 2006; Fynbo et al. 2006; Gal-Yam et al. 2006). Indeed, there is little parameter space of known supernovae that can accommodate these events, suggesting that a genuinely different mechanism may be at play.

There are essentially two strong contenders for this alternative mechanism. One is that in these events, no significant outward supernova shock was launched during the stellar collapse, and black hole formation was direct. Although this conflicts with observations of other GRBs, as noted above, the collapsar model, as first envisaged, anticipated the total collapse of the star during the event that led to the launch of the GRB jets. There is some evidence that might favor this scenario; first, there is an apparent lack of very massive progenitors to supernovae, perhaps suggesting that

more massive stars undergo direct collapse (Smartt 2015); second, there is now evidence that at least one massive star may simply have vanished in exactly this route (Adams et al. 2017). If this is correct, the major question would then appear to be what distinguishes the progenitors of GRBs that launch powerful outward shocks from those that collapse inward.

The second possibility is that these bursts represent a genuinely different physical mechanism from those at play in other long GRBs. For example, they may arise from binary mergers, as suggested for the short GRBs. There is some evidence to support this in the prompt and early afterglow emission of GRB 060614 (Gehrels et al. 2006). It appears to be a short burst with extended emission. The brightest period of prompt emission is short, has a luminosity consistent with that of short GRBs, and also apparently shows zero spectral lag, which measured the difference in time between features visible in different energy bands. Long GRBs typically have a positive lag–luminosity relation, so that softer emission arrives later than harder emission. Short GRBs have no lag, and this is what is observed for GRB 060614. Information on GRB 060505 is more difficult to come by, as it was not promptly triggered on the spacecraft. However, it is also of relatively short duration (~ 4 s) and so may be related to a merging object.

Unfortunately, and perhaps remarkably, other events with comparable constraints on the presence of supernovae have not been found, partly because of the redshift constraints involved (i.e., most bursts are more distant). There are a handful of other bursts for which the evidence of supernovae comparable in luminosity to SN 1998bw is weaker, but this may be due to poorer follow-up, as bursts have become more frequent and follow-up more selective.

5.2 Observational Constraints on Stellar Masses and Sizes

The connection of GRBs exclusively with hydrogen-poor supernovae directly implies hydrogen-poor progenitors (or that any remaining hydrogen can effectively be hidden in the explosion). Hence, stars such as the red supergiants that have been directly observed as the progenitors of hydrogen-rich Type II supernovae in the local universe (Smartt 2009) are unlikely to be promising GRB progenitors. Instead, the most likely progenitor systems that can be located in the local universe are the Wolf–Rayet stars that have shed their envelopes either in strong stellar winds or via interaction with a binary companion. Single Wolf–Rayet stars are expected to start their evolution with very high masses but ultimately lose much of this mass (the entire hydrogen and sometimes helium envelope) to strong stellar winds. They end their lives as compact stars with masses of perhaps $10 M_{\odot}$ (on explosion) and sizes of a few solar radii.

There are few constraints available on the size of GRB progenitor stars. The radius of the emitting region of the GRB itself is extremely large, since the GRB is generated through interactions between shocks several hundred s after they were ejected from the progenitor, and the afterglow is only visible once the outgoing shock has decelerated and lost a significant fraction of its energy. Hence, observations of individual GRBs themselves provide relatively little information as to the

size of the progenitors. However, statistically, it is possible to infer some constraints from the duration distribution of the bursts. The duration of a burst is the time for which the jet is active after it has pierced its progenitor star ($t_{90} \sim t_{\text{GRB}} = t_j - t_b$), where t_j is the time the jet is powered and t_b is the time to breakout from the progenitor, which is roughly $t_b \sim R_*/c$. Interestingly, when investigating the duration distribution of GRBs, there is a marked plateau at a duration of a few seconds, suggesting that bursts are typically not visible if the duration is less than this. This has been interpreted as an absence of bursts with durations $\ll t_b$ and would therefore directly imply that GRB progenitors are compact with radii of a few light-seconds ($\sim 10^9$ m; Bromberg et al. 2012).

In the handful of low luminosities that have been interpreted as shock breakouts, the radii can be estimated, since the spectra appear to be thermal.¹ However, in these cases, the sizes appear much larger ($\sim 10^{11}$ m $\approx 100 R_\odot$), although it should be noted that this is an emission radius rather than, necessarily, the radius of the star. Indeed, in these cases, the radii are typically hundreds of solar radii, more akin to red giants, despite the absence of any hydrogen in the supernova spectra (Campana et al. 2006). This may be due to some dense circumstellar material surrounding the progenitor. Indeed, it has been suggested that the presence (or absence) of this material may explain the difference between low- and high-luminosity GRBs (Nakar 2015). In any case, its presence in cases where radii can be inferred from the presence of blackbody components does not offer great insight into the radii of the progenitors themselves.

There is also little direct evidence for the progenitor masses, although they can be inferred from a variety of different methods. First, the supernovae themselves provide measurements of the ejecta mass and nickel mass involved in the supernova. From this, it is possible to place some limits on the progenitor masses at the time of the explosion. For example, at the most simple level, the star must have had a mass equal to the ejecta mass + remnant compact object mass (which must be estimated). In practice, far more sophisticated modeling of the supernova spectra are used to derive the estimated properties of the ejecta (Mazzali et al. 2007a, 2007b) and link these back to stellar models of the progenitor stars. In many cases, these support initially very massive progenitors (Mazzali et al. 2003).

Alternatively, the age of the stellar population that created the GRB can be used as a proxy for the mass of the progenitor. In particular, because of the strong dependence of core temperature on mass and of nuclear burning rates on core temperature, more massive stars are much more luminous, $L \propto M^3$ (see Yusof et al. 2013 for a more accurate description), and so burn through their fuel reserve more quickly. In particular, since they also have a larger fuel reserve, their lifetimes, $t \propto M^{-2}$. Hence, if the ages of the progenitors can be estimated, these can be directly converted to stellar masses. It is clear that long-GRB host galaxies frequently contain very young populations, with an average age of ~ 5 Myr, corresponding to very massive stellar progenitors ($> 30 M_\odot$; Levesque et al. 2010). This picture is potentially confused by using the integrated properties of long-GRB hosts that may

¹ For a blackbody, if the luminosity and temperature are known, the radius can be calculated from the Stefan–Boltzmann law ($R = \sqrt{L/4\pi\sigma T^4}$).

contain several stellar populations and are not necessarily dominated by the population that spawned the GRB. Ideally, spatially resolved work would be used. In the handful of cases where this is possible, it supports the young ages and hence high masses of the GRB progenitors (Kröhler et al. 2017).

In the absence of detailed spectroscopy of GRB environments, the locations of the GRBs within their host galaxies are also strong diagnostics because of the strong relationship of luminosity with mass for stars. If GRBs are formed from the most massive stars, then these stars will emit much more light than older, less massive stars, and we will observe these stars immediately around the GRB. Therefore, bursts from the most massive stars should be highly concentrated on the light of their host galaxies. This is exactly what is found for long GRBs, which are far more likely to arise from the brightest regions of their hosts than core-collapse supernovae (Fruchter et al. 2006; Svensson et al. 2010; Blanchard et al. 2017), consistent with progenitor masses $>40 M_{\odot}$ (Larsson et al. 2007; Raskin et al. 2008).

5.3 Other Populations of Long-duration GRBs

Although long GRBs are very numerous and the majority of them appear to be rather typical bursts, the sheer number of systems observed means that the parameter space in terms of luminosity and duration has been mapped far more completely for long GRBs than for other classes of burst. Perhaps unsurprisingly, this has resulted in the identification of systems that do not appear to fit comfortably within the bulk population. This may suggest genuinely new classes of GRB spawned from very different progenitors or indicate something about the extremes of the population, probing the limits of stars that can create GRB-like explosions.

5.4 Low-luminosity GRBs

GRB 980425 was an unusually low-luminosity GRB, and other prime examples of the GRB–supernova connection are also notable for their low luminosities. Low-luminosity bursts have a vastly restricted horizon compared to the high-luminosity examples. For example, a burst with an $E_{\text{iso}} \sim 10^{48} \text{ erg s}^{-1}$ is visible to a distance a thousand times less than a burst with $E_{\text{iso}} \sim 10^{54} \text{ erg s}^{-1}$ for a given detector.² Because of these very different volumes of detection, it is likely that the astrophysical (volumetric) rate of low-luminosity GRBs actually vastly exceeds that of the high-luminosity bursts. Indeed, estimates suggest that low-luminosity events are probably a factor of 100–1000 more common (depending on the relative degree of beaming in both high- and low-luminosity bursts).

A crucial question in the study of these events is whether they are directly related to the long GRBs but simply lie at the faint end of a luminosity function or whether they represent a different population, perhaps drawn from a different physical process entirely. The former argument would seem to be supported by the fact that both populations manifest as GRBs and because of the similarity of the supernovae

² Assuming that the burst light-curve shape, spectrum, etc. is otherwise identical, such that the same trigger criteria would be met.

seen in both high- and low-luminosity events, suggesting that the progenitor stars are the same. In this case, then, the range of luminosities might be interpreted in a variety of ways. For example, it is plausible that the viewing angle compared to the relativistic jet could impact the observed properties of the GRB. Since the solid angle around the jet is much larger than that contained within it, that would naturally explain why low-luminosity bursts would be much more common than their high-luminosity counterparts. However, the details of the expected emission in this scenario depend sensitively on the assumed structure of the jets, and indeed, simple “top-hat” jet models have essentially no prompt emission visible to observers outside their cone. The low-luminosity bursts may also reflect an intrinsic distribution in the luminosity of GRBs—for example, because of their energy reservoir or jet power—scenarios that are also plausible.

Alternatively, the low-luminosity GRBs may represent very different physical phenomena all together. Although GRB 980425 appears to be a rather typical GRB, the majority of the low-luminosity events, of which the prototypes may be GRBs 060218, 100316D, and 171205A, are characterized by emission of very long duration (>1000 s) that is spectrally soft and often contains a thermal component. Instead, it has been suggested that the low-luminosity GRBs may then arise from shock breakout, in which the gamma rays and X-rays that last for thousands of seconds are caused by a shock wave from the core collapse that propagates through the star. This model has had considerable success in describing the observed multiwavelength emission in these cases. Indeed, it is possible that these events represent cases where, rather than launching a successful highly relativistic jet to large radii, the jet fails to fully escape the star (a so-called choked jet).

All scenarios remain under active consideration at this time and have arguments in their favor (as well as significant issues). The remarkable similarity of the supernovae in the two scenarios would be puzzling if they were not somehow related, since broad-lined Type Ic supernovae are particularly rare. Furthermore, late-time radio observations that perform calorimetry and determine the total energy budgets suggest that total energies in low- and high-luminosity GRBs are actually rather similar, despite the very different gamma-ray components. Alternatively, it has been suggested that while some low-luminosity bursts have a very long duration, others are short, perhaps shorter than the breakout time of the collapsar, casting doubt as to their connection (Bromberg et al. 2011). Other attempts to model the bulk properties of the population and compute their luminosity functions conclude that it is difficult to remedy the low and high luminosities with typically expected luminosity functions (for example, power-law or broken power-law distributions) and imply that there is a lack of bursts with luminosities intermediate between the high- and low-luminosity populations (Virgili et al. 2009).

5.5 Extremely Long Gamma-Ray Transients

The duration distribution of long GRBs, as observed by BATSE, appears to peter out at durations of around $t_{90} \sim 1000$ s, and few bursts have durations beyond this. However, BATSE was in a relatively low Earth orbit, having been launched from

the space shuttle *Atlantis* in 1991. Because of this, Earth occultation blocked its view of a significant fraction of the sky at a given time, and very long bursts may have been interrupted. A similar problem exists for *Swift*. Other detectors—in particular those at large distances from the Earth, such as detectors on board distant probes such as those used by the IPN—do not have these issues; however, they are typically markedly less sensitive to bursts, in particular if longer events have lower peak fluxes. The very long duration end of the GRB distribution has therefore traditionally been poorly explored; although individual examples of very long bursts have been found, these are generally at durations just beyond 1000 s and are not clearly distinct from the bulk of the GRB population.

Swift has offered solutions to this problem. Although it is unable to view a given patch of sky for long periods of time, its BAT instrument is capable of triggering sources in the image domain, as well as just via increases in the gamma-ray count rate. This means that it is capable of identifying sources that have much lower count rates but much longer durations. Indeed, it can reacquire sources through its gamma-ray detectors in subsequent orbits should they be bright enough. This might be the case if, for example, a burst emitted the same energy but over a much longer period. This has led to the identification of a handful of events with extremely long durations. Indeed, *Swift* has a transient monitoring program that reconstructs images of the sky on timescales of several days and can hence locate far longer events than previously possible (Krimm et al. 2013).

At the very longest end is a population of events that are capable of triggering the BAT on multiple occasions. The prototype for this class of event is *Swift* J1644+57, which was discovered as GRB 110328A on 2011 March 28 but ultimately triggered the BAT on four occasions. Such behavior would be rather typical for some classes of galactic binary system, but observations of *Swift* J1644+57 revealed a redshift of $z = 0.35$. At this distance, the peak luminosity was in excess of 10^{48} erg s $^{-1}$, and with a duration of gamma-ray emission in excess of 10⁶ s, the total isotropic energy output appears to exceed all but the most energetic GRBs. Further observations revealed the source to be coincident with the nucleus of the host galaxy (Levan et al. 2011). These observations are perhaps most naturally, although not uniquely, explained by a very different physical mechanism to GRBs, namely the tidal disruption of a star by the central supermassive black hole. Such events have long been expected (Rees 1988); however, the difference between the expectations of those events and *Swift* J1644+57 was the apparent presence of a relativistic jet (Bloom et al. 2011; Zauderer et al. 2011). Therefore, events like *Swift* J1644+57 (and two other examples, *Swift* J2058+0516 (Cenko et al. 2012; Pasham et al. 2015) and *Swift* J1112–8238 (Brown et al. 2015; 2017)) have been ascribed as relativistic tidal disruption events in which, in addition to the hot thermal emission from the accreting material, a relativistic jet is also launched (Bloom et al. 2011; Burrows et al. 2011).

At somewhat shorter durations of around $\sim 10^4$ s exists another population of events, often dubbed ultra-long GRBs (ULGRBs; Levan et al. 2014). Exhibiting gamma-ray durations well in excess of those for other bursts, their t_{90} values are apparently statistically distinct from those of other bursts (Boer et al. 2015; Levan 2015), although

this may be due to selection effects related to the absence of very long-lived but low flux emission in other bursts (Zhang et al. 2014). At first, it was not apparent that the events were related to GRBs, and, as with the original studies of GRBs, both Galactic (Campana et al. 2011) and extragalactic (Thöne et al. 2011) models were proposed. Ultimately, redshift measurements were successful and demonstrated an extragalactic origin (Levan et al. 2014). In this scenario, there is still a range of plausible progenitor models that may create the bursts. First, it is plausible that minor differences in the collapse process in hydrogen-stripped massive stars could result in longer-lived events. These may even be astrophysically common but difficult to observe. In this case, the ULGRBs are really just the extremely long tail of the GRB distribution. Alternatively, the long-lived emission might be related to far longer infall times in the ULGRB progenitors and a resulting longer-lived but lower-luminosity transient. In this scenario, it is appealing to consider the possibility that ULGRBs arise from GRB-like events in very extended stars (as compared to the compact Wolf–Rayet stars favored for most long GRBs; Levan et al. 2014). Blue supergiants have been suggested as the origin in this scenario. Finally, it is also plausible that, like the longest-duration events, the ULGRBs are not related to core collapse but to tidal disruptions. While the shorter timescales do not naturally match those for the disruption of main-sequence stars by supermassive black holes, they do appear in keeping with the dynamical time expected for the disruption of white dwarfs by intermediate-mass black holes.

An import diagnostic of the nature of ULGRBs was provided by in-depth observations of GRB 111209A, which suggested that it could be explained by a magnetar-powered supernova that peaked at a factor of 2–3 brighter than those seen in most GRBs. In this case, the luminosity of the supernova would not be purely driven by the production of radioactive nickel, as is the case for most GRB supernovae, but would arise through the additional input of energy from the magnetar central engine (Greiner et al. 2015). The associated supernova was apparently hydrogen-poor, which may disfavor a supergiant model for the progenitor. However, the spectrum was also blue and largely featureless, so is difficult to tie directly to a specific type of progenitor. Hence, interest in both tidal disruption and supernova models continues for ULGRBs. Indeed, it may ultimately be the case that multiple channels can create these very long duration bursts.

5.6 Constraints for GRB Production

The creation of a GRB itself is nontrivial. The source region must be compact, since variability is seen in millisecond timescales. The material must be accelerated to high Lorentz factors to explain the production of high-energy photons rather than electron–position (or other) pairs. The engine must switch on (and off again) on timescales varying from a fraction of a second to several thousand seconds. As noted in Chapter 4, this leads to central engines and requirements for rotation for the formation of either a black hole with a centrifugally supported disk or a magnetar.

5.7 Binary or Single?

A crucial question regarding the nature of GRB progenitors is whether the stars that create GRBs arise through single-star or binary-star evolution.³ The evolution and ultimate fate of massive single stars over their short (few million yr) lifetimes is primarily dominated by their initial mass, rotation, and metallicity. In principle, more massive stars at the zero-age main sequence should yield more massive stars immediately prior to supernova. However, the rotation and metallicity dramatically impact this final state. Increased metal content within the stellar atmosphere yields much a higher opacity for outgoing photons due to the many electrons (and associated transitions) in heavier elements. Hence, the outgoing photons have a higher chance of an interaction, creating an excited atom and imparting outward momentum. While the excited electron will naturally revert to the ground state, reemitting a photon of equal wavelength (in the rest frame of the atom) to that of the incoming photon that caused the excitation, this outgoing emission is isotropic, while the incoming photons have a preferred direction radially outward from the star. Hence, this results in the bulk motion of material outward. Since the atom is now traveling outward, the energies of its transitions are redshifted, so it can then absorb outgoing radiation at slightly different frequencies. This process repeats many times, creating strong radiatively driven winds, the strength of which is strongly correlated with the metal content of the star. Hence, as metallicity increases, stars lose progressively more of their mass on the main sequence. Since the energy production within massive stars is strongly dependent on the mass of the star (the bolometric luminosity is commonly assumed to scale as M^3 , although the true scaling is slightly more complex; Yusof et al. 2013), initially more massive stars of the same metallicity may drive stronger stellar winds. Therefore, for stars at high metallicity, the initially most massive stars ($>100 M_{\odot}$) end their lives with masses substantially lower than those that started with lower masses ($\sim 10 M_{\odot}$). Indeed, at high metallicity, black hole formation may be very difficult, if not impossible, due to the effects of this mass loss. This is strikingly demonstrated in some young star-forming regions within the Milky Way, where the most massive stars still visible have masses in excess of $40 M_{\odot}$ but the supernovae to date have created neutron stars and not black holes. For example, the Galactic star cluster G10.0-0.3 contains the magnetar (and hence neutron star) SGR 1806-20 but contains main sequence stars with masses around $50 M_{\odot}$, implying that this was the likely progenitor mass for the star that created the magnetar (Bibby et al. 2008). In order to have created the neutron star, it must have lost a significant fraction of its initial mass before supernova.

The stellar winds that result in large-scale mass loss from stars at high metallicity also have major implications for the rotation of the stars. The material expelled by the stellar winds is carried to large radii but remains coupled to the star via its

³ Most massive stars are formed in binary stars, so the question of interest here is not whether the star is in a binary or not but whether the presence of the binary companion is crucial to the evolution. For example, in some wide binaries, the companion is not important because it never interacts with the other star; alternatively, in other binaries, strong interactions between the two stars are very important.

magnetic field. This carries angular momentum away from the star, reducing its rotation rate. This issue is further exacerbated since, as stars evolve, they enter a red-giant phase, when the star expands to much larger radii. Since the core remains coupled to the envelope during this phase, further braking occurs, resulting in a slowly rotating core. At high metallicity, even stars that begin their lives with very high rotation rates (close to the so-called breakup rate, where the star's rotation can overcome its self-gravity) will lose sufficient angular momentum to end their lives with little rotation and will be unable to create any disk material to power a GRB during their collapse.

It may therefore appear that single stars are unlikely to drive GRBs. However, there is good evidence that GRBs are drawn preferentially from stars at low metallicity. In particular, their host galaxies appear to generally have metallicities of $<1/3$ solar, and this is backed up by detailed analysis of the metallicities of the local environments of the bursts in local cases where such work can be undertaken. At these low metallicities, stellar winds are much less problematic, both in terms of reducing the stellar mass during the main sequence and in carrying away angular momentum. Low-metallicity single massive stars may therefore be a promising route to the creation of at least some GRBs.

A further possibility that has received significant attention in recent years is that massive, low-metallicity, and initially rapidly rotating stars may undergo so-called chemically homogeneous evolution. The principle of such evolution is that in rapidly rotating stars, large-scale circulations of material (often called Eddington–Sweet circulations) can occur, distributing material from the stellar core to regions further out in the star and resulting in the mixing of enriched (i.e., matter that has undergone nucleosynthesis in the core of the star) and nonenriched material. In this model, the star then remains smaller and has no giant phase. Since there is no strong core–envelope evolution, there is little breaking of the stellar core, and ultimately a more massive core can be formed. The collapse of this rapidly rotating core is then a promising location for the creation of a GRB.

The majority of very massive stars are found in binaries, and in a significant fraction of these systems, the two stars will interact, potentially transferring mass and angular momentum between the two stars and dramatically altering their evolution. Indeed, even in the absence of the direct transfer of mass, the impact of the tidal forces on close stars can trigger a response—for example, the circulations that lead to chemically homogeneous evolution. The addition of a binary companion adds a range of additional parameters that impact the final fate of a star. In addition to mass, metallicity, and rotation of the primary star, the same parameters of the secondary also come into play, as does the separation of the binary. The addition of these binary evolution channels therefore greatly increases the diversity of the end products of stellar evolution. Since the majority of massive stars are in such systems, the final outcomes of all massive stars are dominated by the binary routes. Indeed, some 70% will exchange mass, and one-third will create a merged, more massive binary (Sana et al. 2012).

These binary evolution routes are potentially productive routes for the creation of GRB progenitors. The interactions themselves are complicated but can be split into

three broad classes: stripping of the envelope, spin-up of the star, and mergers (Sana et al. 2012). Each of these could be valuable for creating the necessary conditions for a GRB. A stripped star loses its hydrogen envelope such that it will explode as a hydrogen-poor supernova, and stars that are spun up via mass transfer gain angular momentum and are more likely to be spinning rapidly at the end of their lives. Mergers can create more massive, also rapidly rotating stars. Hence, at first sight, it would appear that there are many more routes to the creation of GRBs via binary channels than via single-star evolution. However, while the appeal is obvious, the details of this evolution are less clear. For example, when material is stripped from a star, it will carry away angular momentum, such that the stripped star may not be rotating sufficiently rapidly. The star that accretes material and is spun up has in fact accreted H-rich material and thus may be rapidly rotating but may not create a stripped envelope system. Hence, while binaries clearly can create the necessary conditions for GRB creation, it is still only in a small subset of binary systems that these channels are likely to work effectively.

Indeed, different specific pathways to binary progenitors have been considered. One possibility is to transfer mass (and angular momentum) to the star late in its evolution, potentially reforming a massive, rapidly rotating core, even at high metallicity, where earlier mass loss has reduced both the mass and rotation of the core. Mass transfer during or after helium burning may produce the necessary conditions for GRB formation at solar metallicity and beyond (Brown & Lee 2004). If the mass transfer is unstable (for example, because the Roche lobe shrinks more rapidly than the star as it loses mass), then a runaway event can occur in which much of the envelope mass is transferred rapidly. Since this cannot all be accreted by the secondary star, the result is a so-called common envelope, in which the envelope of the initially mass-transferring star covers both components of the binary, which orbit inside it. Importantly, the envelope is not rotating with the binary, so the stars experience significant dynamical friction as they move through the envelope. Because of this, energy and angular momentum are removed from the binary orbit and used to unbind the common envelope. The binary may be driven to merge in this scenario, but if it is not (i.e., the envelope is unbound before the components merge), then the resulting system is a much tighter binary than at the onset of mass transfer. In this case, tidal forces may synchronize the rotation of the stars with the orbit in so-called tidal locking. For a sufficiently tight binary, this will result in the conditions for angular momentum within the core being met (Lee et al. 2002). Such tidal locking can, in principle, occur with any binary, should the separation be sufficiently tight, and has been considered in a variety of situations (Izzard et al. 2004; Levan et al. 2006; van den Heuvel & Yoon 2007). However, the extent to which these simple models recreate the conditions in nature is unclear, since the core may uncouple from the envelope or winds beyond the common envelope could still drive the binary further apart (and lower core angular momentum). Indeed, simulations of tidal locking fail to reproduce the necessary conditions for GRB creation in tidally locked binaries at solar metallicity (Detmers et al. 2008).

It is also relevant to consider that mergers could create the necessary conditions for GRBs. Since roughly 1/4 of binary stars ultimately merge (and a higher fraction

start with an orbital period of a few days; Figure 5.3), mergers could prove to be a powerful way of creating the necessary angular momentum. Indeed, the merger of two similar-mass helium cores creating a very massive core that rapidly undergoes a supernova has been proposed as a GRB mechanism. Similarly, the merger of a helium core with a black hole has been suggested as a possible route to GRB production, perhaps explaining the ULGRBs (Thöne et al. 2011).

One interesting feature of binary routes to the creation of GRBs is the sheer number of possibilities that arise. Given the broad luminosity function of GRBs, this may not be surprising, and it would be natural to expect that there would be a large range of stellar properties at the point of core collapse. However, the GRB itself is primarily governed by the energy input and shocks, so in itself, it provides little information about the progenitor. On the other hand, supernovae enable the ejecta mass and nickel yield to be calculated and can, with detailed stellar models, potentially be traced back to the properties of their progenitors. In this case, the relatively smaller diversity in GRB supernova properties might be surprising if it indeed reflects a very wide range of progenitor properties. This could suggest that if binary routes are important, then only a small fraction of them create transients akin to GRBs. However, in this case, the challenge may be in creating progenitors at the rates necessary to explain the observed GRB population.

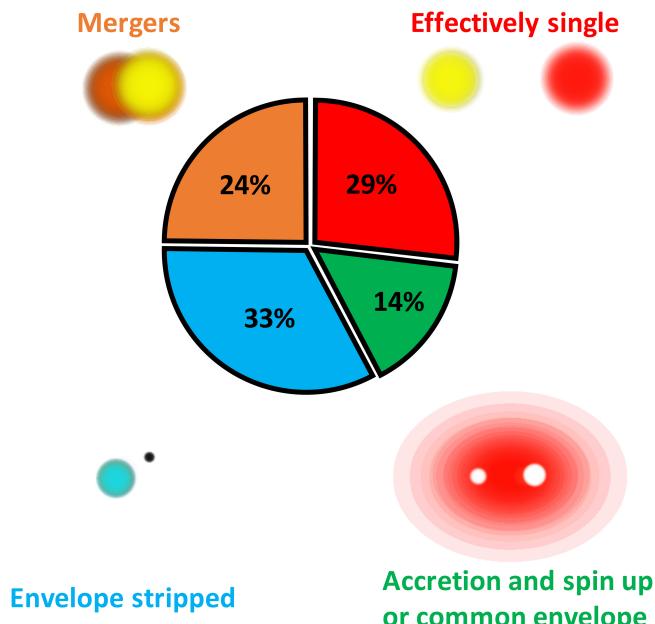


Figure 5.3. Outcomes of binary evolution for massive O stars (numbers from Sana et al. 2012). This shows the overall outcomes, including those that evolve effectively as single stars, in which the envelope is stripped, that are spun up, and that merge. Only a minority of relatively wide binary stars evolve as if they are single, so the majority of stars in binaries (which is the vast majority of all stars at these masses) are impacted significantly by their binarity.

References

- Adams, S. M., Kochanek, C. S., Gerke, J. R., Stanek, K. Z., & Dai, X. 2017, *MNRAS*, **468**, 4968
- Bibby, J. L., Crowther, P. A., Furness, J. P., & Clark, J. S. 2008, *MNRAS*, **386**, L23
- Blanchard, P. K., Berger, E., Fong, W., et al. 2017, *ApJL*, **848**, L22
- Bloom, J. S., Giannios, D., Metzger, B. D., et al. 2011, *Sci*, **333**, 203
- Bloom, J. S., Kulkarni, S. R., Djorgovski, S. G., et al. 1999, *Natur*, **401**, 453
- Bodenheimer, P., & Woosley, S. E. 1983, *ApJ*, **269**, 281
- Boer, M., Gendre, B., & Stratta, G. 2015, *ApJ*, **800**, 16
- Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2011, *ApJ*, **740**, 100
- Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2012, *ApJ*, **749**, 110
- Brown, G. C., Levan, A. J., Stanway, E. R., et al. 2017, *MNRAS*, **472**, 4469
- Brown, G. C., Levan, A. J., Stanway, E. R., et al. 2015, *MNRAS*, **452**, 4297
- Brown, G. E., & Lee, C.-H. 2004, *NewA*, **9**, 225
- Burrows, D. N., Kennea, J. A., Ghisellini, G., et al. 2011, *Natur*, **476**, 421
- Campana, S., Lodato, G., D'Avanzo, P., et al. 2011, *Natur*, **480**, 69
- Campana, S., Mangano, V., Blustin, A. J., et al. 2006, *Natur*, **442**, 1008
- Cano, Z. 2014, *ApJ*, **794**, 121
- Cano, Z., Wang, S.-Q., Dai, Z.-G., & Wu, X.-F. 2017, *AdAst*, **2017**, 8929054
- Cenko, S. B., Krimm, H. A., Horesh, A., et al. 2012, *ApJ*, **753**, 77
- Della Valle, M., Chincarini, G., Panagia, N., et al. 2006, *Natur*, **444**, 1050
- Della Valle, M., Malesani, D., Benetti, S., et al. 2003, *A&A*, **406**, L33
- Detmers, R. G., Langer, N., Podsiadlowski, P., & Izzard, R. G. 2008, *A&A*, **484**, 831
- Fruchter, A. S., Levan, A. J., Strolger, L., et al. 2006, *Natur*, **441**, 463
- Fynbo, J. P. U., Watson, D., Thone, C. C., et al. 2006, *Natur*, **444**, 1047
- Gal-Yam, A. 2012, *Sci*, **337**, 927
- Gal-Yam, A., Fox, D. B., Price, P. A., et al. 2006, *Natur*, **444**, 1053
- Galama, T. J., Tanvir, N., Vreeswijk, P. M., et al. 2000, *ApJ*, **536**, 185
- Galama, T. J., Vreeswijk, P. M., van Paradijs, J., et al. 1998, *Natur*, **395**, 670
- Garnavich, P. M., Stanek, K. Z., Wyrzykowski, L., et al. 2003, *ApJ*, **582**, 924
- Gehrels, N., Norris, J. P., Barthelmy, S. D., et al. 2006, *Natur*, **444**, 1044
- Greiner, J., Mazzali, P. A., Kann, D. A., et al. 2015, *Natur*, **523**, 189
- Hjorth, J., Sollerman, J., Møller, P., et al. 2003, *Natur*, **423**, 847
- Izzard, R. G., Ramirez-Ruiz, E., & Tout, C. A. 2004, *MNRAS*, **348**, 1215
- Jakobsson, P., Hjorth, J., Malesani, D., et al. 2012, *ApJ*, **752**, 62
- Jakobsson, P., Levan, A., Fynbo, J. P. U., et al. 2006, *A&A*, **447**, 897
- Kouveliotou, C., Woosley, S. E., Patel, S. K., et al. 2004, *ApJ*, **608**, 872
- Krimm, H. A., Holland, S. T., Corbet, R. H. D., et al. 2013, *ApJS*, **209**, 14
- Kröhler, T., Kuncarayakti, H., Schady, P., et al. 2017, *A&A*, **602**, A85
- Larsson, J., Levan, A. J., Davies, M. B., & Fruchter, A. S. 2007, *MNRAS*, **376**, 1285
- Lee, C.-H., Brown, G. E., & Wijers, R. A. M. J. 2002, *ApJ*, **575**, 996
- Levan, A., Crowther, P., de Grijs, R., et al. 2016, *SSRv*, **202**, 33
- Levan, A. J. 2015, *JHEAp*, **7**, 44
- Levan, A. J., Davies, M. B., & King, A. R. 2006, *MNRAS*, **372**, 1351
- Levan, A. J., Tanvir, N. R., Cenko, S. B., et al. 2011, *Sci*, **333**, 199
- Levan, A. J., Tanvir, N. R., Starling, R. L. C., et al. 2014, *ApJ*, **781**, 13
- Levesque, E. M., Berger, E., Kewley, L. J., & Bagley, M. M. 2010, *AJ*, **139**, 694

- Lipkin, Y. M., Ofek, E. O., Gal-Yam, A., et al. 2004, *ApJ*, **606**, 381
- Mazzali, P. A., Deng, J., Tominaga, N., et al. 2003, *ApJL*, **599**, L95
- Mazzali, P. A., Foley, R. J., Deng, J., et al. 2007a, *ApJ*, **661**, 892
- Mazzali, P. A., Kawabata, K. S., Maeda, K., et al. 2007b, *ApJ*, **670**, 592
- Nakar, E. 2015, *ApJ*, **807**, 172
- Pasham, D. R., Cenko, S. B., Levan, A. J., et al. 2015, *ApJ*, **805**, 68
- Pian, E., Amati, L., Antonelli, L. A., et al. 2000, *ApJ*, **536**, 778
- Raskin, C., Scannapieco, E., Rhoads, J., & Della Valle, M. 2008, *ApJ*, **689**, 358
- Rees, M. J. 1988, *Natur*, **333**, 523
- Sana, H., de Mink, S. E., de Koter, A., et al. 2012, *Sci*, **337**, 444
- Smartt, S. J. 2009, *ARA&A*, **47**, 63
- Smartt, S. J. 2015, *PASA*, **32**, e016
- Soderberg, A. M., Chakraborti, S., Pignata, G., et al. 2010, *Natur*, **463**, 513
- Svensson, K. M., Levan, A. J., Tanvir, N. R., Fruchter, A. S., & Strolger, L.-G. 2010, *MNRAS*, **405**, 57
- Thöne, C. C., de Ugarte Postigo, A., Fryer, C. L., et al. 2011, *Natur*, **480**, 72
- van den Heuvel, E. P. J., & Yoon, S.-C. 2007, *Ap&SS*, **311**, 177
- Virgili, F. J., Liang, E.-W., & Zhang, B. 2009, *MNRAS*, **392**, 91
- Woosley, S. E. 1993, *ApJ*, **405**, 273
- Yusof, N., Hirschi, R., Meynet, G., et al. 2013, *MNRAS*, **433**, 1114
- Zauderer, B. A., Berger, E., Soderberg, A. M., et al. 2011, *Natur*, **476**, 425
- Zhang, B.-B., Zhang, B., Murase, K., Connaughton, V., & Briggs, M. S. 2014, *ApJ*, **787**, 66

Gamma-Ray Bursts

Andrew Levan

Chapter 6

Short-GRB Progenitors

6.1 Introduction

Short GRBs are the population of bursts generally defined as having prompt gamma-ray durations of $t_{90} < 2$ s and an on average harder spectrum than their longer cousins. They have been identified as a separate subpopulation since the early 1980s (Mazets and Golenetskii 1981), although it was not until the first results from BATSE were available that such a distinction became statistically strong (Kouveliotou et al. 1993). The duration distribution from BATSE and *Swift* is shown in Figure 6.1, and the existence of a second population with shorter durations is clear (although it is also clear that at $t_{90} = 2$ s, there is a significant overlap of the two populations).

Despite intense interest, studies of the afterglows of short GRBs lagged almost a decade behind those of long bursts. In retrospect, this is not surprising; while the peak flux of short GRBs is typically comparable to that of long GRBs, their shorter durations significantly reduce their total fluence. It is now apparent that there is a broad correlation between the fluence of a GRB and the brightness of its afterglow (albeit with significant scatter), so it unsurprising that short-GRB afterglows are typically fainter than those of long GRBs (Gehrels et al. 2008; Nysewander et al. 2009). Combined with their rarity, it was therefore extremely challenging to locate them. Indeed, while several afterglow searches were conducted for short GRBs in the late 1990s and early 2000s, none of these were successful (Gorosabel et al. 2002; Hurley et al. 2002b). Part of this problem was related to the low rate of bursts discovered in this era. As with *Swift*, its forerunners, *BeppoSAX* and *HETE-2*, were optimized for a rather softer response than BATSE and thus detected few short bursts. Because of this, it was not until 2005 that the first afterglows were discovered with the X-ray afterglow of the *Swift* GRB 050509B (Gehrels et al. 2005; Bloom et al. 2006) and the optical afterglow of GRB 050709 (Fox et al. 2005; Hjorth et al. 2005), a burst actually detected by *HETE-2*.

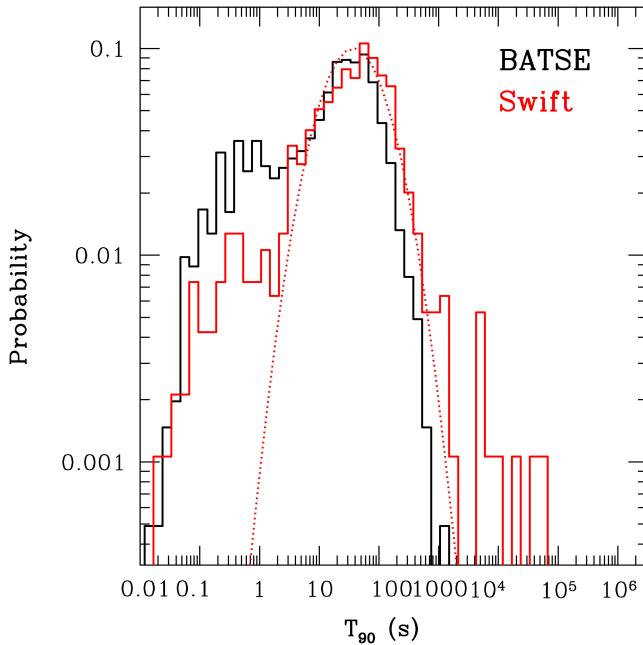


Figure 6.1. Duration distribution of GRBs detected by the *Swift*-BAT and BATSE (from Levan 2015). A lognormal distribution is a reasonable description for the long GRBs (dashed red line), but a clear population of short bursts is also visible (see also Chapter 2). Short GRBs represent $\sim 25\%$ of the bursts detected by BATSE but only $\sim 10\%$ of the bursts detected by *Swift*. This is likely due to the difference in detector response, with the *Swift* detectors preferentially detecting softer photons and thus long GRBs.

As with the long bursts, the detection of afterglows enabled the positions of short GRBs to be pinpointed on the sky, their redshifts to be measured from their host galaxies, and their environments and locations to be studied in detail. There are now several tens of short GRBs for which X-ray and/or optical afterglow measurements are available, and from these, it has been possible to piece together a consistent picture of their creation. More recently, this has been greatly aided by the detection of faint, radioactively powered transients (kilonovae; Berger et al. 2013; Tanvir et al. 2013) and even the measurements of gravitational waves in coincidence with a short GRB (Abbott et al. 2017b). This chapter outlines the likely progenitor models discussed for short-duration GRBs and how the observations of afterglows and host galaxies compare to the expectations of these most popular models.

6.2 Progenitor Models

Early in the afterglow era, once it was clear that GRBs were cosmological, the models for the short GRBs were largely the same as considered for the long bursts, namely either the collapse of a massive star (Woosley 1993) or the merger of two compact objects (Li and Paczyński 1998; Paczyński 1998). While the assertion that short GRBs are related to merging binaries had been significantly made once the long GRBs were firmly associated with massive star collapse, there are other reasons

to consider short GRBs as more promising candidates for binary mergers. In particular, the merger itself is naturally very rapid: even for black hole–neutron star mergers, the merger is over in a few orbits and thus is unlikely to last more than a few seconds (if that). In contrast, in massive star models, the act of core collapse and the subsequent infall of material can take place on far longer timescales. Further, the duration of a GRB is given by the lifetime of the central engine after the jet has broken out of whatever material surrounds the compact object that is launching the jet (Bromberg et al. 2013). In the case of long GRBs, this means that the jet, traveling at relativistic velocities, is probably crossing the star on a timescale of a few seconds (but importantly, clearly longer than 2 s) and then is visible as a GRB for as long as it continues to be powered, having broken out. For it to last long enough to break out but <2 s beyond this point would require a good degree of fine-tuning. On the other hand, for binary mergers, there is very little material that the jet must puncture, so shorter durations do not create any uncomfortable tensions.

6.2.1 Creation of DCO Binaries

Creation in the Field

The majority of double compact object (DCO) binaries are probably created in the field through various binary evolution channels. However, this creation is not a straightforward process. In particular, the binary must remain bound through the evolution, notwithstanding the effects of mass loss at the point of each supernova, in addition to the natal kicks given to neutron stars (Fryer et al. 1998; Arzoumanian et al. 2002) and probably black holes (Repetto et al. 2012). Finally, to be interesting as a short GRB progenitor, it must finish this process with a sufficiently small separation to merge within a Hubble time. These separations are a few solar radii, often orders of magnitude smaller than the physical size of each progenitor star at certain points in the evolution of the system. In practice, there is no single route to this progenitor creation, and detailed stellar evolution calculations, as well as rapid population synthesis (Portegies Zwart & Yungelson 1998; Belczynski et al. 2002; Hurley et al. 2002a), find various routes in which massive binaries can create DCO systems. However, there are certain features common to many of these routes, and a “standard” picture is generally accepted and shown graphically in Figure 6.2. In this scenario, two stars, both of which are above the mass limit for supernova formation,¹ are found in a binary. They evolve separately (as single stars) through the main-sequence lifetime of the initially more massive star (the so-called primary). This star ultimately leaves the main sequence and evolves into a giant, at which time there may be mass transfer onto the companion (often called the secondary). The initially more massive star then normally completes its evolution and explodes as a supernova, leaving a compact object and a main-sequence secondary star, with its evolution possibly altered by the accretion of material. This star then continues to

¹ Formally, since mass transfer from the more massive to less massive star occurs, it is possible for one of the stars to be less massive or, indeed, in extremum, for the initially more massive star to lose so much mass that it ultimately becomes a white dwarf (Church et al. 2006).

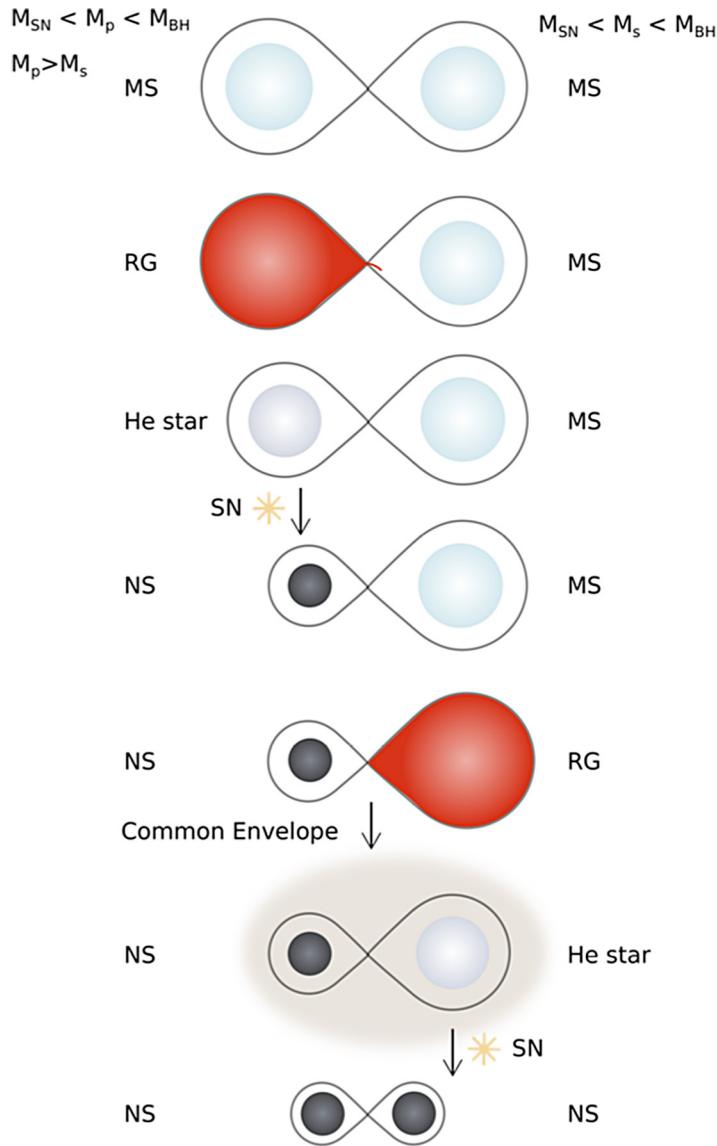


Figure 6.2. Typical route to the creation of a double neutron star binary (see also Fryer et al. 1999). Image courtesy Rachel Tunnicliffe. Two massive stars (which will ultimately form neutron stars) are initially in an orbit well within their respective Roche lobes. As the initially more massive star evolves, it fills its Roche lobe and may transfer mass to the secondary. It will then evolve to become a bare He core and explode as a supernova. If the binary survives this, then when the second star evolves to become a giant, it fills its Roche lobe, leading to unstable mass transfer and the creation of a common envelope. This envelope extracts orbital angular momentum and energy while it is unbound, and the resulting binary is a neutron star–neutron star system that will merge within a Hubble time.

evolve and ends its main-sequence life. At the point at which it expands to become a giant, it begins to transfer mass onto the neutron star. This mass transfer is unstable and ends with the formation of a common envelope: both stars orbiting within the extended atmosphere of one star. The resulting He core and neutron star system finds itself in a dense environment in which dynamical friction extracts angular momentum and energy from the orbit, dramatically tightening the binary (shortening its period) while unbinding the remaining stellar envelope. At the end of this process, the envelope is gone and the binary has shrunk, perhaps from separations on scales of astronomical units down to scales of solar radii. The secondary star can then explode as a supernova, leaving a binary consisting of two neutron stars.

The scenario is generally similar for systems containing black holes, although the required masses in this case are typically higher, so the rates of formation are generally expected to be lower. Indeed, while approximately 10 neutron star–neutron star binaries are known within the Milky Way, there are so far no black hole–neutron star systems. While this may be due in part to their rarity, it is also likely due to the detection mechanism, in which neutron star–neutron star binaries are located via radio emission from a pulsar that may have been spun up during the mass transfer process. Since black holes cannot be detected via this method, their observational nondetection to date is less constraining.

At even higher masses, it is possible to form black hole–black hole systems similar to those seen with advanced gravitational-wave detectors. These are unlikely to be of interest as short GRB progenitors, since the general consensus is that little, if any, electromagnetic light will escape, although there has been a claim of a short GRB from a binary black hole merger (see Chapter 9 and Connaughton et al. 2016, 2018). Nonetheless, they may be formed via similar channels and perhaps preferentially at lower metallicity, where massive stars retain more of their mass and may form massive black holes (de Mink & Mandel 2016; Eldridge & Stanway 2016).

As noted above, the picture described here is necessarily limited. There are many more ways to create such systems in the field, with the precise evolution of a given system dictated by the masses, spins, metallicities, and initial separations of the different components. There may be multiple periods of mass transfer and common-envelope evolution (Belczynski et al. 2002) or, in some cases, more exotic forms of stellar evolution, such as chemically homogeneous evolution (de Mink & Mandel 2016). The details of this evolution mean that attempts to constrain the rate of formation of DCO binaries from theoretical population synthesis remain plagued with significant systematic uncertainties.

Dynamical Creation

While most DCO systems are likely to be formed in the field, it is also possible to create them via interactions in dense regions, such as galactic nuclei or globular clusters. Within such clusters, the time between interactions of stars (especially those in binaries where the cross section is the separation of the binary rather than the size of the individual star) is much less than the Hubble time, so the majority of stars in binaries may experience an interaction (Davies et al. 1992). Interactions between binary and single (or binary and binary) stars tend to leave the most massive stars as

components of binaries, while the less massive stars are ejected. Since most neutron stars have a mass of $\sim 1.4 M_{\odot}$ and the turnoff mass in a globular cluster is $\sim 0.6 M_{\odot}$, this implies that any neutron stars within a globular cluster will be in binary systems. Successive interactions tend to harden the binary, bringing it closer and potentially leading to a merger time much less than the Hubble time. Of the neutron star binaries known within the Milky Way, there is one example that exists in a globular cluster, B2217+11C in M15 (Anderson et al. 1990), that was probably created via this route. It has been suggested that dynamical formation could be a major route of creating neutron star mergers (Grindlay et al. 2006), and indeed, these suggestions also apply for more massive remnants, such as black hole mergers. However, these estimates still contain significant uncertainty. For example, the velocity dispersion in globular clusters is low, and so is the escape velocity. This means that any modest natal kick to a neutron star on formation should eject the newly formed neutron star from its parent cluster. Given estimates of the neutron star kick distribution (from pulsars), it is perhaps surprising that globular clusters retain as many neutron stars as they do, and indeed, the total number of massive remnants within a cluster is clearly a dominant source of error in estimating the total rate of mergers created through such a channel.

6.2.2 Mass Loss and Natal Kicks

During the formation of the DCO binary, various dynamical effects act to determine the final parameters of the system. In particular, in addition to the determination of the remnant masses, there is also the final semimajor axis and eccentricity of the binary and its space velocity relative to birth environment. All of these are impacted by mass loss from the system and natal kicks to the neutron stars and black holes.

Virial equilibrium tells us that the kinetic energy is half the gravitational potential energy within the binary prior to any supernova ($K = -U/2$). In other words, for two stars of initial mass M_1 and M_2 (and total mass $M_T = M_1 + M_2$), we can write

$$E_1 = -\frac{GM_1M_2}{2R_1} = \frac{\mu_1 v_1^2}{2}, \quad (6.1)$$

where μ_1 is the reduced mass of the system ($\mu_1 = M_1M_2 / (M_1 + M_2)$). This means that the velocity of the system is given by

$$v_1 = \sqrt{\frac{G(M_1 + M_2)}{R_1}}. \quad (6.2)$$

At the point of the supernova, this condition applies, but at this time, there is an effective instantaneous mass loss from the supernova-creating star, say $M_1 \rightarrow M_3$ or the change in mass $\Delta M = M_1 - M_3$. However, the velocity at the point of explosion is set by the pre-explosion conditions and does not change immediately afterward (i.e., right at the point of the explosion, the separation is still R_1 and the velocity v_1). This means that since the mass of the system has reduced, the remnant with mass M_3

will now be moving too quickly for the new binary and will move to a larger orbit, R_2 . The energy post-explosion is then

$$E_2 = -\frac{GM_3M_2}{2R_2} = -\frac{GM_3M_2}{R_1} + \frac{\mu_2 v_1^2}{2}. \quad (6.3)$$

If $E < 0$, then the orbit remains bound, and the case where $E = 0$ is the one where the binary is (just) unbound, or R_2 tends to infinity. In this limit,

$$-\frac{GM_3M_2}{R_1} + \frac{\mu_2 v_1^2}{2} = 0, \quad (6.4)$$

or, substituting for the reduced mass, this simplifies to

$$\frac{G(M_T - \Delta M)}{R_1} = \frac{v_1^2}{2} = \frac{GM_T}{2R_1}. \quad (6.5)$$

This yields a solution when $\Delta M = M_T / 2$. In other words, if more than half of the total mass of the binary is lost in the supernova, then the system will be unbound.

This appears problematic. Imagine, for example, an evolution of two massive stars with initial masses of $M_1 = 12$ and $M_2 = 10 M_\odot$. The initially more massive star will evolve more rapidly and explode as a supernova, leaving behind a neutron star of mass $M_3 = 1.4 M_\odot$. In this case, and with the simple assumption that the mass is lost instantaneously, the binary will survive, since $(M_1 - M_3)/(M_1 + M_2) = 10.6/22 = 0.47$. However, the binary will then continue to evolve toward the second supernova. At the point of this explosion, it will form another neutron star of mass $M_4 = 1.4 M_\odot$, losing a further $8.6 M_\odot$ in the process. The mass-loss ratio at this point is $(M_2 - M_4)/(M_3 + M_4) = 8.6 / 11.4 = 0.75$, and the binary will be unbound.

In practice, many binaries are unbound at the point of supernova, and this is one reason why the DCO merger rate is orders of magnitude below the core-collapse supernova rate, despite many massive stars arising in binary systems that will ultimately interact. However, clearly some neutron star–neutron star binaries are formed, as they are observed in the field. This is due to two effects. The first is that mass loss is actually gradual. The above example assumes that mass is lost as an impulsive and effectively instantaneous event. In practice, stellar winds or binary interactions often carry away much of the mass of the system well before the supernova. Indeed, the core of a star immediately before the supernova explosion is often only modestly (a factor of <3) more massive than the neutron star it creates, especially at high metallicity. If the mass is lost gradually, then the effect is to widen the orbit but not to unbind it. Indeed, in close binaries, binary interactions can in fact shrink the orbit while mass is being lost, for example, in a common envelope.

Nonetheless, mass lost during a supernova explosion can result in the unbinding of the binary. Even if it does not, conservation of momentum will result in a kick to the remaining binary. The mass lost does so at a certain point in the orbit and thus has a preferred direction. The binary then recoils such that momentum is conserved. Therefore, the binaries will not remain in their birthplace but may attain space velocities of tens to hundreds of km s⁻¹.

Furthermore, it is clear from the proper motions of pulsars that neutron stars can receive significant kicks on formation that are independent of the binary (Arzoumanian et al. 2002). These so-called natal kicks (kicks imparted to neutron stars at birth) can be several hundred km s⁻¹. They can act to provide additional eccentricity to the newly formed binary and also unbind systems that would otherwise have remained bound or, conversely, keep bound systems that should have been unbound. The origin of neutron star kicks remains uncertain, with various models ranging from simple differences in infalling mass onto the proto–neutron star due to deviations from spherical symmetry in the progenitor star to complex sound waves created during the supernova process. Since the origin of these kicks is uncertain, it is less clear that the black hole population receives such kicks, although the locations of some black hole X-ray binaries well above (or below) the galactic plane do favor a kicked scenario (Repetto et al. 2012).

This combination of mass loss and kicks typically means that a binary that does survive to create a DCO will be highly eccentric at formation and have a space velocity that is comparable to those of isolated neutron stars. This velocity is likely to be similar to the escape velocity of the Galaxy in which the system was born, so some DCO binaries will escape directly to the intergalactic medium while others remain bound. Those that do remain bound will be placed on eccentric orbits determined by the gravitational potential of the Galaxy in which they orbit. They will spend most of their time at large galactic radii and are most likely to merge well away from their birth sites, or the cores of their host galaxies.

6.2.3 Gravitational-wave Radiation-driven Mergers

From the point at which the second neutron star is formed, the evolution of the binary itself is relatively straightforward to compute (much simpler than the evolution that led to its formation). The semimajor axis and eccentricity will evolve following the analytic prescription first provided by Peters (1964), namely

$$\frac{da}{dt} = -\frac{64}{5} \frac{G^3}{c^5} \frac{m_1 m_2 (m_1 + m_2)}{a^3 (1 - e^2)^{7/2}} \left(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4\right) \quad (6.6)$$

for semimajor axis and

$$\frac{de}{dt} = -\frac{304}{15} \frac{G^3}{c^5} \frac{e m_1 m_2 (m_1 + m_2)}{a^4 (1 - e^2)^{5/2}} \left(1 + \frac{121}{304} e^2\right) \quad (6.7)$$

for eccentricity. The central feature to note in these equations is the strong dependence on both the separations and eccentricity, with the merger time scaling as a^4 . Systems that are either extremely tight or highly eccentric will merge much more rapidly. Unless the initial eccentricity is extreme, the majority of systems will circularize prior to the final merger. For these circular systems, the equations above can be readily integrated to give the merger time as a function of the semimajor axis as

$$t_{\text{merge}} = \frac{5}{256} \frac{c^5}{G^3} \frac{a^4}{m_1 m_2 (m_1 + m_2)}, \quad (6.8)$$

providing a straightforward route to determining merger times (Mandel & de Mink 2016) from the formation of the binary neutron star. Note that the delay from star formation may be slightly more complex to compute, since this includes the time for the stellar evolution to take place, likely several million years and potentially longer.

During the final stages of the merger, the semimajor axis and hence orbital period are changing rapidly, and the system will take only a few minutes to move from ~ 10 Hz to merger at ~ 1 kHz. During this time, the gravitational-wave emission is sufficiently strong (strain amplitude sufficiently high) that the gravitational waves can be directly detected by ground-based interferometers such as LIGO and VIRGO (Abbott et al. 2016b), as has now been done for several black hole–black hole binaries (Abbott et al. 2016a) and one neutron star–neutron star binary (Abbott et al. 2017b); see Chapter 9 for further details.

The merger itself creates either a rapidly spinning supermassive neutron star or a black hole. In the former case, the magnetic fields may be extreme, and a magnetar may be formed. Depending on the mass of the neutron star and the neutron star equation of state, the newly created neutron star may remain stable or be supported against collapse by rotation, in which case it can collapse to a black hole as it is slowed by its own magnetic field.

In essence, the conditions within the central engine are therefore very similar to those created in the collapse of a massive star to form a long-duration GRB, in that an accreting black hole or highly magnetized neutron star is formed. This central engine goes on to power the GRB emission via either neutrino–antineutrino annihilation (Rosswog et al. 2003) or magnetic acceleration. See Chapter 4 for more details.

6.2.4 Properties at Formation and Delay-time Distribution

One crucial question about the population of compact object binaries created is what the expected delay-time distribution is. In other words, how long does it take each binary to merge after formation? This is a complex problem to address, since the merger time scales as a^4 , such that a small range of semimajor axis makes a very large difference in the merger time. It is commonly assumed that the distribution of merger times is flat in $\log t$ (i.e., equal numbers of mergers per decade in time), but this is not necessarily the case. One crucial question relates to the presence (or absence) of a population of binaries with very short merger times ($< 10^7$ yr) that are comparable to the stellar evolution time to create the binaries.

Some evidence has been found suggesting that these short delays could be important or event-dominant (Belczynski & Kalogera 2001) because of a pathway to formation in which the first formed neutron star is not recycled and where both stars enter the helium giant phase together, creating a double common envelope (Belczynski & Kalogera 2001). It would be difficult to confirm the existence of such a short-lived population within the Milky Way because those young systems would

merge rapidly and then be undetectable; see Figure 6.3 (Belczynski et al. 2006). This could be important because a large, unseen population of binaries would dramatically increase the astrophysical rate of compact object mergers and hence of short GRBs (if associated) and gravitational-wave transients. It would also make the use of environmental properties, such as the age of the underlying stellar population, a much more challenging tool to distinguish between different progenitor channels, since binaries with very short delay times may have locations indistinguishable from core-collapse supernovae.

6.3 Prompt Emission Properties

As the name suggests, short GRBs are defined as those with a short duration of prompt emission. The division between short and long GRBs is typically set at 2 s, based on the apparent saddle in two overlapping lognormal distributions that provided a good description of the duration distribution of BATSE GRBs (see Figure 6.1). Within the population of short GRBs, the same broad range of properties observed for long GRBs can be seen. In particular, there is the same great diversity in the morphology of the short emission, with single sharp peaks and

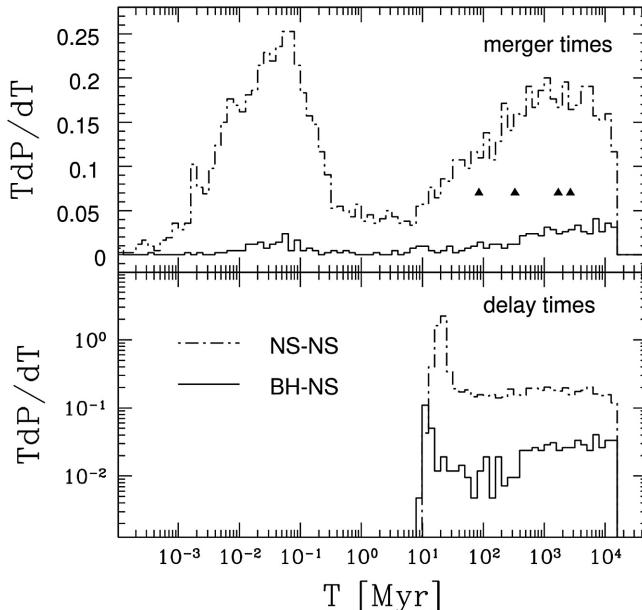


Figure 6.3. Delay-time distribution of black hole–neutron star and neutron star–neutron star binaries (from Belczynski et al. 2006). The top panel shows the distribution of merger times due to gravitational-wave radiation (from the formation of the second neutron star or black hole), while the lower panel shows the range of delay times from the onset of star formation (i.e., from the point at which the binary formed on the zero-age main sequence). The filled triangles show the merger times of known galactic neutron star–neutron star binaries that will merge within a Hubble time. It is notable that in this prediction, there is a significant population of events with very short (<1 Myr) merger times. If correct, these are an important population that is difficult to observationally confirm within the Milky Way because the majority of such systems have already merged. © 2006. The American Astronomical Society. All rights reserved.

multiple emission episodes observed. The distribution of peak flux for the short GRBs is very similar to that of the long bursts, but due to the much shorter emission periods, their integrated fluence over the burst period is typically smaller.

Spectrally, short GRBs are harder than long GRBs, emitting more high-energy photons (Kouveliotou et al. 1993). They also appear to show zero lag between the harder and softer emission, in contrast to long GRBs, which do show a spectral lag that correlates with the luminosity of the burst (Gehrels et al. 2006). The different spectra of the two populations imply that missions with differing sensitivities and spectral passbands will observe the bursts to be different. This is an obvious statement in terms of the relative rates of short and long GRBs— instruments with better soft sensitivity will observe more long bursts, and systems with better hard sensitivity will observe relatively more short bursts. Because of this, *Swift* observes 10% of its triggers as short bursts, while BATSE observed \sim 25%. In raw numbers, this means that for $t_{90} = 2$ s, the contribution of the long population is larger when observed with *Swift* than when observed with BATSE. However, it is more complex than this, because the impact of the spectrum, sensitivity of the instrument, and spectral lags can mean that the burst duration is itself a function of the energy at which the duration is measured. Thus, the defining line between long and short GRBs, where there is, say, a 50% probability of a burst being in a given class, is actually not at a fixed duration but at a mission- (or energy- and sensitivity-) dependent duration. This has been investigated by Bromberg et al. (2013), who considered spectral hardness to be a more important discriminator than duration. In particular, they determined the contribution of collapsars at short duration based on the plateau in the cumulative duration distribution introduced by the breakout time from the collapsar emission, providing a different route to estimating the contribution of collapsars within the short-GRB population than simply measuring the duration itself. From this, they determined that because of the softer response of *Swift*, it detects bursts mainly at the softer end of the hardness duration distribution. Hence, the dividing line, at which a burst has equal probability of being either short or long, lies at 0.7 s for *Swift*-detected bursts.

This approach does have some drawbacks. In particular, it is based on rather simple spectral models (for example, not considering the broken power laws that actually provide a better fit to the broadband spectra of GRBs). It also relies on a physical interpretation of an apparent plateau in the duration distribution as being due to the breakout time of long GRBs from their progenitor stars. Nonetheless, it clearly highlights that identifying a burst as a short GRB is significantly more complex than simply identifying a burst with a duration of less than 2 s. Indeed, to have high confidence in a burst belonging to the short-burst category, it is necessary for it to be both very short (ideally, less than 1 or even 0.5 s) and spectrally hard.

However, even this approach is imperfect, since it has become apparent based on observations with *Swift* that a further population of short bursts exists in which extended emission is also present (often these bursts are referred to as SGRB+EE bursts). In these cases, the measured t_{90} can be greatly in excess of 2 s, but the burst itself appears to belong in the short-GRB category. A strong example of this is GRB 050724 (Barthelmy et al. 2005; Berger et al. 2005). The burst has a measured t_{90} of

96 s but an initially short spike. It is well localized by an optical afterglow to an elliptical galaxy and thus clearly does belong in the short-GRB category. Similarly, the supernova-less GRB 060614 (Della Valle et al. 2006; Fynbo et al. 2006; Gal-Yam et al. 2006) has also been claimed to belong in this class (Gehrels et al. 2006), implying that new routes are needed to cleanly identify short GRBs from long ones.

However, this task is not trivial. It is appealing to use the properties of the afterglow or the host galaxy to make this decision (Levan et al. 2007), since, for example, an elliptical galaxy is unlikely to host a long GRB. This leads to the suggestion that alternative classification mechanisms can be used that, rather than linking to duration, instead link to the physical progenitor mechanism. This has been trialed by Zhang et al. (2009), who suggested Type I and Type II GRBs. Type I GRBs are caused by binary mergers (or whatever mechanism creates unambiguously short GRBs), and Type II events are collapsars (or the mechanism that creates unambiguously long GRBs). This is potentially powerful but risks overinterpreting the early observational data and could still readily misclassify some bursts, for example, a short GRB+EE in a highly star-forming galaxy.

6.4 Afterglow Properties

While afterglows were discovered for long GRBs in 1997, the first short-GRB afterglow was not uncovered until 2005 via the X-ray afterglow of GRB 050509B. The faintness of the afterglow emission is likely related to the low fluence observed for prompt emission, as studies have shown a broad correlation in this regard (Gehrels et al. 2008; Nysewander et al. 2009). In turn, this yields fainter X-ray and optical afterglows. While X-ray afterglows are still commonly discovered for short GRBs, they are by no means certain. *Swift* observes X-ray afterglows from around 70% of short GRBs, whereas almost all long GRBs exhibit X-ray afterglow emission. In the optical, afterglows are uncovered for the majority of short GRBs with fluence of $>10^{-7}$ erg cm $^{-2}$, at least where reasonably deep searches have been undertaken within a few hours of the burst itself, but they are not found for fainter bursts (Nysewander et al. 2009). Interestingly, the population of bursts where the optical is clearly suppressed relative to the expectations from X-ray observations (the so-called dark bursts) is much smaller in the short-burst population. This may reflect differing progenitor pathways in which short GRBs are not exploding in dense, dusty environments, although part of this is likely because the faint X-ray afterglows require exceptionally deep optical observations to be constraining on the X-ray-to-optical spectral slope.

Once identified, monitoring of the afterglows of short GRBs shows them to be generally consistent with the fireball model determined for long GRBs, although again, the paucity of data means that constraints are naturally weaker. The evolution of the fireball is broadly as expected, but importantly, the long-term monitoring enables a reconstruction of the particle number density (n) around the progenitor. In cases where a measurement of this has been possible, it appears that particle number densities for short GRBs are indeed much lower than the average for long bursts, with $n = 3\text{--}15 \times 10^{-3}$ cm $^{-3}$ (Fong et al. 2015). This again may represent the very different locations in which long and short GRBs are found.

Given the faintness and the difficulty tracking the afterglows to late times, it is perhaps unsurprising that garnering evidence regarding collimation in short GRBs has proved challenging. Measurements of the jet break, when $\theta_j \sim 1/\Gamma$, where θ_j is the jet opening angle and Γ is the Lorentz factor, require the identification of an achromatic break in the afterglow decay. Ideally, this requires tracking both the X-ray and optical afterglow to late times. This has been possible in a few cases, although it frequently requires the combination of *Chandra* with large ground-based telescopes. A handful of claimed jet breaks with varying degrees of robustness have been claimed in the literature, and these provide opening angles typically in the region of $15^\circ \pm 10^\circ$ (Fong et al. 2015).

6.5 Host Galaxy Properties

Host galaxy properties can be particularly diagnostic of progenitor systems because they can provide a handle on the ages of the stellar populations responsible for production of the transient. Well before core-collapse and thermonuclear supernovae were clearly separated by their progenitors (massive star and white dwarf, respectively), it was clear that their origins must be different because Type II supernovae occurred exclusively in star-forming galaxies, while hydrogen-deficient (in particular Type Ia) supernovae occurred in both young and ancient stellar populations. Since massive stars have short lives, it was therefore apparent that Type Ia supernovae were not exclusively originating from young progenitors. This now fits cleanly with the model we have for their production, in which they are created by carbon–oxygen white dwarfs being driven over the Chandrasekhar mass.

Short- and long-GRB host galaxies bear marked similarities to the dichotomy between Types II and Ia supernovae. The host galaxies of long GRBs are exclusively star-forming, often at quite high rates when considering the low stellar mass of the system (they have a high specific star formation = star formation rate per unit stellar mass). The picture is very different for short GRBs; GRB 050509B apparently arose from a passive and extremely massive elliptical galaxy (Gehrels et al. 2005; Bloom et al. 2006), and while this is not common, it is clearly true that some short bursts can arise from elliptical systems and others from more typical spiral galaxies in which the star formation mode is more akin to that of the Milky Way (Fong et al. 2013). The breakdown of galaxy types for short GRBs is shown in Figure 6.4. Although the details of the galaxy-type fractions depend somewhat on how the sample is constructed (for example, using just precise optical positions or also those with X-ray–only positions, etc.), a broad picture emerges in which roughly half of the bursts arise from late-type galaxies and just 10% from early-type systems (see Figure 6.5). There also exists a population of bursts for which the available data make it difficult to type the host galaxies and a further population for which there is no high-confidence host galaxy, the so-called “hostless” bursts (Berger 2010; Tunnicliffe et al. 2014; see below).

The galaxy-type fractions, as well as the detailed properties of the galaxies themselves, potentially provide information about the delay-time distributions for the progenitors. For progenitors with very long delay times, it would be reasonable

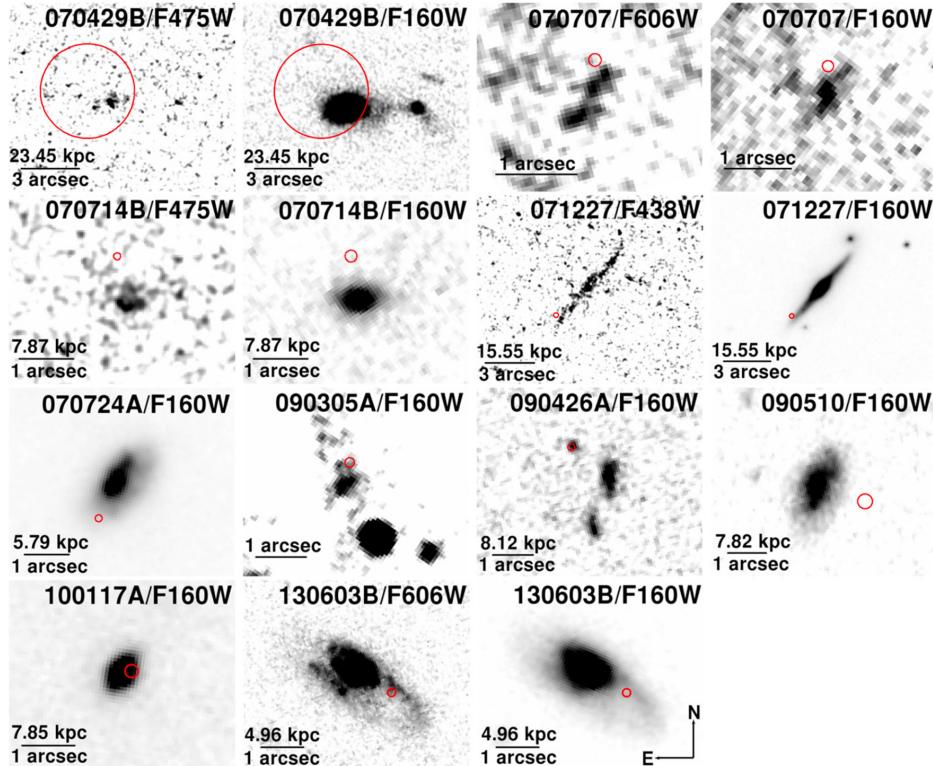


Figure 6.4. Sample of high-confidence short-GRB host galaxies observed in multiple filters with *HST*, from Fong and Berger (2013). © 2013. The American Astronomical Society. All rights reserved.

to expect the rate of short GRBs to trace the stellar mass of the galaxies, whereas for very short delays, it should trace the star formation rate. Formally, the rate will actually follow the star formation rate convolved with the delay-time distribution, which is likely very broad; however, in the extremum of very long or very short delays, it will closely follow either the mass or star formation rate. In the local universe, around 20% of the stellar mass is found in early-type elliptical galaxies (so-called T types of -6 to -4), a further 70% is found in spirals (S0 to Sd or T types -3 to 7), and 10% is found in irregulars (de Vaucouleurs et al. 1991; Levan et al. 2006). In contrast, for star formation, the fractions are 0%, 80%, and 20%. Interpreting these numbers is tricky given the presence of populations of short GRBs that are hostless and for which the host type cannot be readily identified, but it seems plausible that the mergers are tracing the stellar mass, which suggests a long delay time. In this case, bursts that arise in star-forming hosts will not, in general, be drawn from the youngest populations but from preexisting older populations of stars within the galaxies.

This picture can also be tested on individual host galaxies, although there are significant difficulties in extracting reliable star formation histories from the very limited data generally available. Instead, it is more often possible to isolate a

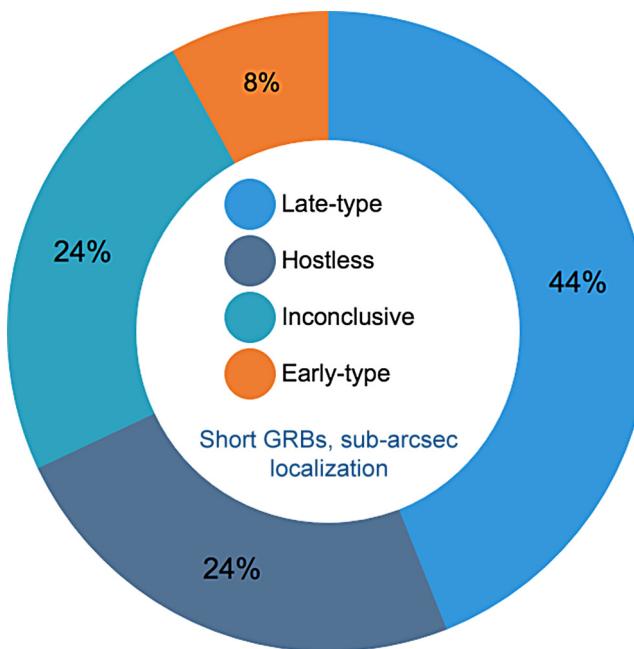


Figure 6.5. Galaxy-type fractions of well-localized short GRBs (data from Fong et al. 2013). Given the redshifts of some bursts, there is a significant fraction of host galaxies for which the galaxy type is uncertain. There also exists a population of apparently hostless bursts. These may arise from progenitors kicked from more local galaxies or residing in more distant and undetected galaxies. Only a small fraction ($\sim 8\%$) of short GRBs arise from clearly early-type systems.

characteristic age for the dominant stellar population within the Galaxy. For short GRBs, these are generally old, at several hundred million yr (Leibler & Berger 2010), well beyond the evolution time for massive stars. However, in some cases, younger populations are clearly (and unsurprisingly) present and may not be related to the burst progenitor.

In addition to these properties, there is also a striking population, which includes up to 25% of the short bursts for which no host is identified in deep imaging, the so-called hostless bursts (Berger 2010; Tunnicliffe et al. 2014). In these bursts, there is no underlying galaxy, nor is there any obvious galaxy proximate on the sky. Attempting associations based on the probability of locating galaxies at different distances does not yield any very strong candidates. In some cases, even deep observations with *HST* fail to uncover the host galaxy. There are different possibilities for the origin of these bursts. They may represent events at extreme offsets from their host galaxies, either because of kicks to the binary or because of dynamical formation in (remote) globular clusters. Given that the latter scenario would require 25% of all GRBs to be formed in such remote clusters, and such clusters only represent a small fraction of the total cluster population, this scenario seems unlikely. Alternatively, it is possible that the bursts represent a high-redshift population for which the host galaxies are unseen. Indeed, at high redshift, many

long-GRB hosts have magnitudes fainter than $R \sim 26$, where the limits for hostless short bursts are often set. Locating distant hosts of short GRBs may in fact be more difficult than for long bursts, since long-GRB hosts are bright rest-frame UV emitters that can be observed in the optical window at $z > 2\text{--}3$. However, if short-GRB hosts are dominated by older populations even at this epoch, then it is plausible that fainter galaxies could be missed. Ultimately, even deeper observations, perhaps with the *James Webb Space Telescope*, may be necessary to firmly distinguish between these possibilities. Indeed, it is possible or even likely that both scenarios are represented in nature, and it is just the relative fractions of hostless bursts that are kicked and at high redshift that must be determined.

6.6 Locations

For DCO mergers, the space velocities of hundreds of km s^{-1} propel the binaries from their birth sites. The expected final locations are a function not only of the binary properties (in particular the expected merger time and velocity) but also of the location of their birth and the properties of the host galaxy. In particular, the location within the galaxy, combined with the galaxy mass, sets the effective escape velocity from that location. The galaxy mass (stellar + dark matter) determines the galactic potential in which the binary moves. More massive galaxies have higher escape velocities and thus will retain more of their binaries, probably in highly elliptical orbits in which binaries spend the most time at larger radii. In contrast, lower-mass galaxies will have lower escape velocities, and the binaries will be unbound and escape to large distances.

On a smaller scale, the merger time will also impact the locations. Short merger times ($\sim 10^6\text{--}10^7$ yr) will result in binaries that merge close to their formation sites and at times when young stars or other signs of star formation may be recognized.² Longer merger times will result in far larger distances traveled by the binary, although the motion through the galaxy potential will mean that this does not directly correspond to the distance traveled, since it may have experienced several bound orbits of its host. For extreme delay times, the assumption of an effectively static galaxy is almost certainly an oversimplification, since galaxies undergo interactions as part of hierarchical structure formation and thus the potential will change dramatically over billions of years. Since the binaries are often only loosely bound to their hosts, these interactions may have quite profound effects on the expected locations of DCO binaries.

Observationally, it is, in principle, straightforward to measure the offset of a short GRB from its host galaxy, and it only becomes nontrivial in cases where either the error in the location of the burst is large or the offset is so large that the host galaxy cannot be straightforwardly identified. It is also possible to measure a so-called normalized offset, in which the offset is expressed in units of the host galaxy effective (half-light) radius.

²It is useful to note that a velocity of 1 km s^{-1} corresponds to approximately 1 pc in 1 million yr. Hence, velocities and offsets can be scaled to first order.

These measured offsets are typically large. A short-GRB mean offset is 4.5 kpc, much larger than the ~ 1.5 kpc seen for long GRBs (Fong & Berger 2013). In host-normalized terms, 75% lie beyond the half-light radius of their host galaxies. In a population that followed the stellar population within the host, it would be expected that 50% of bursts would arise from within the half-light radius (cf. long GRBs; Bloom et al. 2002; Blanchard et al. 2017), so these large offsets are evidence for some form of kick to the progenitor systems.

It is also possible to characterize the locations on smaller scales, in particular using the F_{light} parameter (Fruchter et al. 2006). In this case, F_{light} represents the fraction of the total light of the Galaxy that arises from regions with a lower surface brightness than that containing the transient, so that $F_{\text{light}} = 1.0$ is a source on the brightest pixel and $F_{\text{light}} = 0$ lies off the galaxy light (see Chapter 8 for examples with long GRBs). For short GRBs, approximately 50% of the bursts lie off the stellar light of their hosts ($F_{\text{light}} = 0$), a percentage that is only marginally lower in the infrared, which traces older stellar populations (Fong et al. 2013); see Figure 6.6. This in turn suggests that the short GRBs are not merely associated with an older stellar population that is faint, but would be visible in red light, but are indeed kicked from their formation locations.

This picture qualitatively matches the expectations for compact object mergers, since they can have long merger times and kicked progenitors. However, ascertaining the details of this is more complex. In particular, the expected locations are not fixed (e.g., a fixed set of offsets for all galaxies) but are a function of the galaxy type

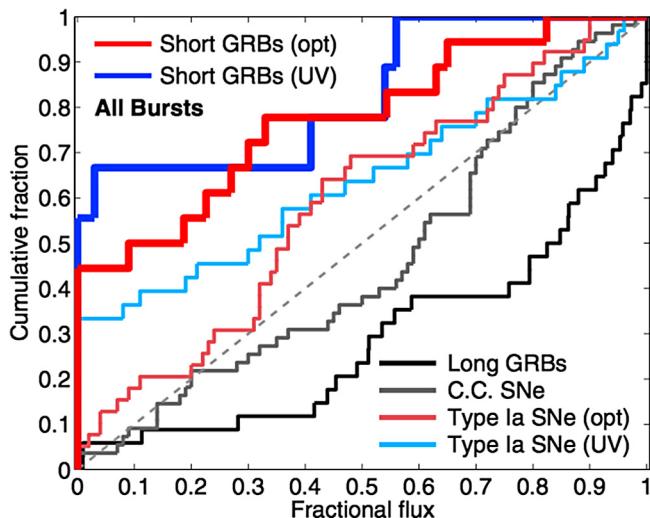


Figure 6.6. Distribution of short GRBs on their host galaxy light (from Fong et al. 2015). In this case, a value of one indicates a burst on the brightest pixel of its host, and a value of zero is a burst that is off the light of its host galaxy. It can clearly be seen that many short GRBs apparently lie well off their host galaxies. This is inconsistent with them being drawn from any population that remains tied to the general stellar population of the hosts and suggests that the burst progenitors have been kicked from their formation sites. © 2015. The American Astronomical Society. All rights reserved.

itself. It is therefore necessary to consider this in asking if the observed locations are consistent with the expectations for compact binary mergers. This has been attempted by using generic galaxy potentials (Fong et al. 2013) or by trying to reconstruct them for individual galaxies (Church et al. 2011).

The general picture from these works is that given the uncertainties associated with the prediction of compact object populations, the bursts are, in general, consistent with these expectations. However, there are some examples that appear to lie at larger distances than expected from very massive galaxies. These binaries are unlikely to be ejected to such locations if the binary evolution creating them is correct. Instead, it is likely that they represent either a population drawn from globular clusters or, alternatively, a set where the host galaxy has been misidentified. This may be likely in some cases, since the bursts at the largest distances from their hosts are exactly those for which the probabilistic association is weakest.

6.7 Redshifts and Energetics

As with the long GRBs, there was a major observational effort to acquire reliable distance indicators for the short GRBs. These distances have proved more challenging, largely because of the faint optical afterglows that have made obtaining absorption spectroscopy particularly difficult. Indeed, the majority of short-GRB redshifts arise from their host galaxies, either in emission or absorption, and there is likely only one direct absorption redshift for a short GRB (de Ugarte Postigo et al. 2014). Since short GRBs are also frequently well offset from their host galaxies, the associations are not always secure and sometimes have a significant probability of chance alignment. However, in the majority of cases, the host galaxies are identified with reasonable certainty and thus can be used to determine burst energetics.

The isotropic energies of short GRBs, as with long bursts, span a very wide range of energy, $10^{49} \text{ erg} < E_{iso} < 10^{52} \text{ erg}$, but are on average almost two orders of magnitude less energetic than long GRBs (Frail et al. 2001; Fong et al. 2015), with a peak in the broad distribution around 10^{51} erg . They also exhibit a similar range in kinetic energy (Fong et al. 2015). As noted above, the opening angles associated with short GRBs are frequently poorly constrained due to the faintness of their afterglows but have an average that is broadly similar to that of the long-GRB population. This means that, corrected for beaming, it is likely that the typical short-GRB energy is in the range of a few $\times 10^{49} \text{ erg}$, again almost a factor of 100 fainter than the long bursts. Given the paucity of jet-break detections, it is less clear how well the application of a beaming correction reduces the scatter in the burst energies.

Given the fainter limits, this also implies a much-limited horizon for short-GRB detection when compared with the long GRBs and that an unseen high-redshift population is more likely, since even the most luminous short GRBs are difficult to detect at high redshift. This is interesting in the context of some very high-redshift GRBs that in their rest frame appear to have durations that would make them appear as short GRBs (i.e., the measured $t_{90} / (1 + z) < 2 \text{ s}$). The absence of lower-redshift analogs suggests that these bursts are unlikely to be drawn from the long-GRB

population and may only appear as “short” bursts because of a combination of detector effects and the application of the $(1 + z)$ correction for cosmological time dilation, which is not normally applied in measuring GRB durations (Littlejohns et al. 2013).

6.8 Radioactively Driven Transients

Perhaps one of the most discussed routes to pinpointing the progenitors of short GRBs was the possibility of observing a longer-lived optical transient powered by the decay of radioactive material synthesized in the ejecta from the neutron star. The merging process is not entirely clean, particularly for mass ratios that deviate from unity, and some material is ejected. This material is neutron-rich (at least to start with) and thus is a likely place for the creation of neutron-rich elements via the so-called *r*-process. The subsequent decay of this material produces a thermal transient that is expected to evolve more rapidly and be rather fainter than a supernova. Hence, the detection of such a transient in a short GRB could be a smoking gun as to its origin, in much the same way as a supernova unmasks the progenitors of long GRBs. In the literature, a radioactively driven transient created in the merger of two neutron stars or a neutron star and a black hole will be referred to variously as a Li–Paczynski mini-supernova (Li & Paczynski 1998), a macronova (Kulkarni 2005), or a kilonova (Metzger & Berger 2012). Here we use the kilonova, but to first order, the three events can be considered the same.

6.8.1 Production of Kilonovae

The principle of kilonova production is that during the merger of either two neutron stars or a neutron star and a black hole, there are various sources of mass loss from the system. Initially, as the binaries are deformed during the merger, some material is ejected in the equatorial plane in the form of tidal tails. A polar dynamical component may also be released at this time. The ejecta is from a neutron star it is initially neutron-rich, which is defined by the electron fraction, Y_e , which is directly equivalent to the proton fraction. The tidal tails are normally expected to be electron-poor, $Y_e \sim 0.1$, while polar ejecta should contain more electrons, $Y_e \sim 0.3$ (Metzger 2017). As the merger continues, neutrinos from the proto–neutron star and a contribution from a disk wind continue to eject material.

6.8.2 Observational Constraints on Kilonovae

Searches for kilonova emission began immediately after the first discovery of short-GRB afterglows in 2005. Optical observations of GRB 050509B were undertaken in the hours to days after the event and reached impressive depths of $R \sim 27$ (Hjorth et al. 2005; Bloom et al. 2006). However, neither an optical afterglow nor a kilonova were found. The redshift of GRB 050509B is uncertain because of the large X-ray error box and the absence of optical emission. However, searches in the afterglows of other bursts (both with and without optical afterglows) also failed to uncover any clear evidence for kilonovae. The limits associated with these observations ruled out

the most optimistic versions of the models but left plenty of parameter space in which the transients could still be uncovered (Metzger & Berger 2012).

A crucial step forward arose with a better consideration of opacity within the neutron-rich ejecta. In particular, the products of nucleosynthesis in the ejecta of binary mergers are very different from those created in supernovae. However, the assumptions regarding radiative transfer of energy through the ejecta and to the observer that were employed were typically those for iron-group elements that are dominant in supernovae. Instead, the r -process elements created in mergers should have a high opacity. In particular, lanthanides have open electron f -shells, creating a very large number of possible transitions across the optical regions. While the opacities of all the lanthanides remain uncertain based complex numerical calculations, the effect is clear. Optical light in lanthanide-rich ejecta should be essentially extinguished. Instead, we would expect to observe a somewhat longer-lived but much redder transient (Barnes & Kasen 2013).

The opportunity to test this hypothesis arose soon after it was first made with the short-burst GRB 130603B at $z = 0.35$ (de Ugarte Postigo et al. 2014). This burst exhibited an optical counterpart that was well mapped from the optical to infrared at early times. Later observations taken with the *HST* revealed a dramatic color change. Ten days after the burst, the source was undetectable in optical light but clearly visible in the infrared, with an absolute magnitude of $M_H \sim -16$. Indeed, the infrared light lay well above the extrapolation of the afterglow and suggested the presence of an additional component. This component was interpreted as a kilonova and matches the expectations from the lanthanide-rich models (Berger et al. 2013; Tanvir et al. 2013); see Figure 6.7. This discovery was broadly in keeping with the models developed by Barnes & Kasen (2013), although other authors suggested that the necessary mass ejected to explain the luminosity, of order $0.1 M_\odot$, was at the upper end of what was likely, perhaps making the kilonova associated with GRB 130603B unusually bright.

Once this discovery was announced, archival searches were made in previous short-GRB data. These were typically not well designed for searching for kilonovae; in particular, infrared observations at later times were frequently not pursued. Beyond this, the issues of distances are even more acute for short GRBs and kilonovae than they are for long GRBs and supernovae. They are a factor of 10–100 fainter and peak in the infrared, where ground-based observations begin to be plagued by a bright sky and atmospheric absorption. Because of this, observations, even from 8 m telescopes, are only sensitive out to $z \sim 0.15$, while *HST* is needed to reach to $z \sim 0.4$. Nonetheless, despite these difficulties, moderately convincing signatures of kilonovae were uncovered in two further GRBs: GRB 050709 ($z = 0.16$; Jin et al. 2016) and GRB 060614 ($z = 0.12$; Jin et al. 2015). The latter example is actually formally a long burst for which no supernova was uncovered (Fynbo et al. 2006) but has also been suggested to arise from a short burst with extended emission (Gehrels et al. 2006). In this case, the kilonova detection may be taken as evidence for the validity of the latter proposal.

Unfortunately, there have been few opportunities to launch comprehensive campaigns to observe kilonova, since few bursts since 2013 have both been at the

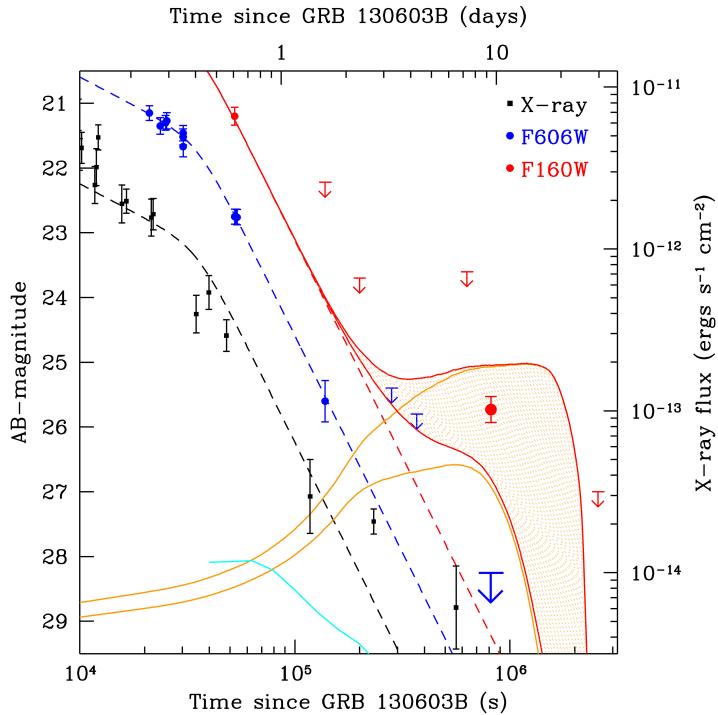


Figure 6.7. The X-ray, optical, and infrared light curve of GRB 130603B (modified from Tanvir et al. 2013). The optical light appears to decay with the same power law observed in the X-rays. However, the infrared does not follow this decay, instead appearing much brighter on a timescale of 10 days postburst. This is well explained by the range of kilonova models proposed using opacities from *r*-process elements, in particular the lanthanides. Reprinted by permission from Springer Nature: Tanvir et al. (2013).

requisite redshift and had rapidly identified optical counterparts. Those events that have been observed, if the redshift assignations are correct, apparently harbor kilonovae fainter than that seen in GRB 130603B (Tanvir et al. 2015).

6.8.3 Heavy-element Production

Perhaps the crucial question that observations of kilonovae can answer is the origin of the heavy, so-called *r*-process elements. These elements, which include gold, silver, iodine, and uranium, as well as all of the lanthanides, are clearly vital to the heavy-element chemical enrichment seen on Earth and, indeed, may even be important for the evolution of life, given the importance of, e.g., iodine for human health. However, identifying the astrophysical sites that are responsible for their production has been plagued with difficulties.

Elements up to iron are naturally synthesized in massive stars as they evolve toward supernovae, and other heavy elements, can be formed in supernova ejecta or potentially via asymptotic giant branch (AGB) stars. The route through which these elements are formed is called the *s*-process. In this scenario, the rate of neutron capture is much smaller than the radioactive decay time, so each neutron capture

creates a more massive isotope, but as soon as an unstable nucleus is produced, it will β -decay to a higher atomic number well before the next neutron capture.³ Relatively low rates of incident neutrons can be found in supernovae ejecta or in the outer regions of AGB stars. Crucially, the *s*-process is restricted to creating elements that can be straightforwardly produced by single β -decays and so lie along the so-called valley of stability.

However, there are also *r*-process elements that are formed via rapid neutron capture. In order to form *r*-process elements, there must be successive neutron captures, which push a given atom to a heavier isotope. In the *r*-process, the capture of neutrons must be faster than the decay time of radioactive isotopes in order to build progressively heavier atoms. Eventually, these will decay, often via a β -decay, which results in the creation of an element with a higher atomic number. The conditions for the *r*-process are extreme, since the neutron fluxes and temperatures must be extremely high. However, the mergers of two neutron stars have long been viewed as a possible source, since there are abundant free neutrons in any material ejected during the merger. Other possibilities also exist, including an origin for these elements in supernovae.

Distinguishing the astrophysical sites responsible for the creation of the *r*-process elements is a major goal of contemporary astrophysics that overlaps many different areas of astrophysics, physics, and geophysics. Indeed, some of the strongest evidence for the origin of the heaviest elements arises from estimates of the rate of events that create the elements. Studies of plutonium on the seafloor (Wallner et al. 2015) that has been accreted from an interstellar flux show that it is not refreshed regularly, as might be expected if it were created in supernovae, since multiple supernovae are expected to have occurred in the vicinity of Earth in the past tens of millions of yr. Instead, it appears likely that the plutonium is produced in events that eject more mass in heavy elements but are intrinsically rarer.

6.9 Gravitational-wave Emission

Ultimately, the most direct way to associate a given mechanism with short-GRB production is to directly observe the behavior of mass within the progenitor via gravitational waves (see Chapter 9 for a full discussion of gravitational waves and recent observations). The emission of these waves occurs from any source that is accelerating in a nonspherically symmetric manner, so merging compact objects are the strongest expected sources. The recent detection of gravitational waves from a neutron star binary known as GW 170817 (Abbott et al. 2017b) was strikingly accompanied just 2 s later by the identification of a short GRB by both *Fermi* and *INTEGRAL* (Abbott et al. 2017a, 2017b; Savchenko et al. 2017). This would appear to immediately confirm the hypothesis that at least some short GRBs are created by compact object mergers, cementing the indirect evidence accrued over the past decade. Indeed, this gamma-ray discovery was rapidly followed by the identification of an optical and near-infrared counterpart in the galaxy NGC 4993 (Arcavi et al. 2017;

³There are in practice multiple decay channels for the isotopes created in the *r*-process and all of these contribute energy to the kilonova.

Coulter et al. 2017; Drout et al. 2017; Lipunov et al. 2017; Soares-Santos et al. 2017; Tanvir et al. 2017) and ultimately by observations of likely off-axis radio, optical, and X-ray afterglows (Hallinan et al. 2017; Margutti et al. 2017; Troja et al. 2017; Lyman et al. 2018). These observations are in keeping with those made of short GRBs since the discovery of their afterglows in 2005. However, the redshift of GW 170817/GRB 170817 in NGC 4993 is only 0.01, a distance of 40 Mpc (Hjorth et al. 2017; Cantiello et al. 2018), so the inferred isotropic energy release of the burst is four orders of magnitude below that seen in the *Swift* short GRBs. It is therefore not yet entirely clear what the relation between the burst seen in GW 170817 and the short GRBs seen by *Swift* is. It may simply be a viewing-angle effect, that the burst was very weak viewed at a large off-axis angle but would have appeared highly luminous to an observer on-axis (Lyman et al. 2017; Margutti et al. 2018). Alternatively, it may be that many compact object mergers do not accelerate ejecta to highly relativistic velocities but instead eject them in an only modestly relativistic cocoon (Mooley et al. 2017). In that case, a short GRB would be a rare outcome of a merger, and the true merger rate would be much higher than that inferred from short GRBs. Ultimately, in-depth observations of further gravitational-wave sources will be required to conclusively address this issue, this is discussed in more detail in Chapter 9.

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, *PhRvX*, **6**, 041015
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, *PhRvL*, **116**, 061102
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *ApJL*, **848**, L13
 Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, *ApJL*, **848**, L12
 Anderson, S. B., Gorham, P. W., Kulkarni, S. R., Prince, T. A., & Wolszczan, A. 1990, *Natur*, **346**, 42
 Arcavi, I., Hosseinzadeh, G., Howell, D. A., et al. 2017, *Natur*, **551**, 64
 Arzoumanian, Z., Chernoff, D. F., & Cordes, J. M. 2002, *ApJ*, **568**, 289
 Barnes, J., & Kasen, D. 2013, *ApJ*, **775**, 18
 Barthelmy, S. D., Chincarini, G., Burrows, D. N., et al. 2005, *Natur*, **438**, 994
 Belczyński, K., & Kalogera, V. 2001, *ApJL*, **550**, L183
 Belczynski, K., Kalogera, V., & Bulik, T. 2002, *ApJ*, **572**, 407
 Belczynski, K., Perna, R., Bulik, T., et al. 2006, *ApJ*, **648**, 1110
 Berger, E. 2010, *ApJ*, **722**, 1946
 Berger, E., Fong, W., & Chornock, R. 2013, *ApJL*, **774**, L23
 Berger, E., Price, P. A., Cenko, S. B., et al. 2005, *Natur*, **438**, 988
 Blanchard, P. K., Berger, E., Fong, W., et al. 2017, *ApJL*, **848**, L22
 Bloom, J. S., Kulkarni, S. R., & Djorgovski, S. G. 2002, *AJ*, **123**, 1111
 Bloom, J. S., Prochaska, J. X., Pooley, D., et al. 2006, *ApJ*, **638**, 354
 Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2013, *ApJ*, **764**, 179
 Cantiello, M., Jensen, J. B., Blakeslee, J. P., et al. 2018, *ApJL*, **854**, L31
 Church, R. P., Bush, S. J., Tout, C. A., & Davies, M. B. 2006, *MNRAS*, **372**, 715
 Church, R. P., Levan, A. J., Davies, M. B., & Tanvir, N. 2011, *MNRAS*, **413**, 2004
 Connaughton, V., Burns, E., Goldstein, A., et al. 2016, *ApJL*, **826**, L6
 Connaughton, V., Burns, E., Goldstein, A., et al. 2018, *ApJL*, **853**, L9

- Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, *Sci*, 358, 1556
- Davies, M. B., Benz, W., & Hills, J. G. 1992, *ApJ*, 401, 246
- de Mink, S. E., & Mandel, I. 2016, *MNRAS*, 460, 3545
- de Ugarte Postigo, A., Thöne, C. C., Rowlinson, A., et al. 2014, *A&A*, 563, A62
- de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G. Jr., et al. 1991, Third Reference Catalogue of Bright Galaxies, Volume I: Explanations and References. Volume II: Data for Galaxies between 0^h and 12^h. Volume III: Data for Galaxies between 12^h and 24^h (New York: Springer)
- Della Valle, M., Chincarini, G., Panagia, N., et al. 2006, *Natur*, 444, 1050
- Drout, M. R., Piro, A. L., Shappee, B. J., et al. 2017, *Sci*, 358, 1570
- Eldridge, J. J., & Stanway, E. R. 2016, *MNRAS*, 462, 3302
- Fong, W., & Berger, E. 2013, *ApJ*, 776, 18
- Fong, W., Berger, E., Chornock, R., et al. 2013, *ApJ*, 769, 56
- Fong, W., Berger, E., Margutti, R., & Zauderer, B. A. 2015, *ApJ*, 815, 102
- Fox, D. B., Frail, D. A., Price, P. A., et al. 2005, *Natur*, 437, 845
- Frail, D. A., Kulkarni, S. R., Sari, R., et al. 2001, *ApJL*, 562, L55
- Fruchter, A. S., Levan, A. J., Strolger, L., et al. 2006, *Natur*, 441, 463
- Fryer, C., Burrows, A., & Benz, W. 1998, *ApJ*, 496, 333
- Fryer, C. L., Woosley, S. E., & Hartmann, D. H. 1999, *ApJ*, 526, 152
- Fynbo, J. P. U., Watson, D., Thöne, C. C., et al. 2006, *Natur*, 444, 1047
- Gal-Yam, A., Fox, D. B., Price, P. A., et al. 2006, *Natur*, 444, 1053
- Gehrels, N., Barthelmy, S. D., Burrows, D. N., et al. 2008, *ApJ*, 689, 1161
- Gehrels, N., Norris, J. P., Barthelmy, S. D., et al. 2006, *Natur*, 444, 1044
- Gehrels, N., Sarazin, C. L., O'Brien, P. T., et al. 2005, *Natur*, 437, 851
- Gorosabel, J., Andersen, M. I., Hjorth, J., et al. 2002, *A&A*, 383, 112
- Grindlay, J., Portegies Zwart, S., & McMillan, S. 2006, *NatPh*, 2, 116
- Hallinan, G., Corsi, A., Mooley, K. P., et al. 2017, *Sci*, 358, 1579
- Hjorth, J., Levan, A. J., Tanvir, N. R., et al. 2017, *ApJL*, 848, L31
- Hjorth, J., Watson, D., Fynbo, J. P. U., et al. 2005, *Natur*, 437, 859
- Hurley, J. R., Tout, C. A., & Pols, O. R. 2002a, *MNRAS*, 329, 897
- Hurley, J. R., Tout, C. A., & Pols, O. R. 2002b, *ApJ*, 567, 447
- Jin, Z.-P., Hotokezaka, K., Li, X., et al. 2016, *NatCo*, 7, 12898
- Jin, Z.-P., Li, X., Cano, Z., et al. 2015, *ApJL*, 811, L22
- Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, *ApJL*, 413, L101
- Kulkarni, S. R. 2005, arXiv:[astro-ph/0510256](#)
- Leibler, C. N., & Berger, E. 2010, *ApJ*, 725, 1202
- Levan, A. J. 2015, *JHEAp*, 7, 44
- Levan, A. J., Davies, M. B., & King, A. R. 2006, *MNRAS*, 372, 1351
- Levan, A. J., Jakobsson, P., Hurkett, C., et al. 2007, *MNRAS*, 378, 1439
- Li, L.-X., & Paczyński, B. 1998, *ApJL*, 507, L59
- Lipunov, V. M., Gorbovskoy, E., Kornilov, V. G., et al. 2017, *ApJL*, 850, L1
- Littlejohns, O. M., Tanvir, N. R., Willingale, R., et al. 2013, *MNRAS*, 436, 3640
- Lyman, J. D., Lamb, G. P., Levan, A. J., et al. 2018, *NatAs*, 2, 751
- Lyman, J. D., Levan, A. J., Tanvir, N. R., et al. 2017, *MNRAS*, 467, 1795
- Mandel, I., & de Mink, S. E. 2016, *MNRAS*, 458, 2634
- Margutti, R., Alexander, K. D., Xie, X., et al. 2018, *ApJ*, 856, L18
- Margutti, R., Berger, E., Fong, W., et al. 2017, *ApJL*, 848, L20

- Mazets, E. P., & Golenetskii, S. V. 1981, *Ap&SS*, **75**, 47
- Metzger, B. D. 2017, *LRR*, **20**, 3
- Metzger, B. D., & Berger, E. 2012, *ApJ*, **746**, 48
- Moley, K. P., Nakar, E., Hotokezaka, K., et al. 2017, *Natur*, **554**, 207
- Nysewander, M., Fruchter, A. S., & Pe'er, A. 2009, *ApJ*, **701**, 824
- Paczyński, B. 1998, *ApJL*, **494**, L45
- Peters, P. C. 1964, *PhRv*, **136**, 1224
- Portegies Zwart, S. F., & Yungelson, L. R. 1998, *A&A*, **332**, 173
- Repetto, S., Davies, M. B., & Sigurdsson, S. 2012, *MNRAS*, **425**, 2799
- Rosswog, S., Ramirez-Ruiz, E., & Davies, M. B. 2003, *MNRAS*, **345**, 1077
- Savchenko, V., Ferrigno, C., Kuulkers, E., et al. 2017, *ApJL*, **848**, L15
- Soares-Santos, M., Holz, D. E., Annis, J., et al. 2017, *ApJL*, **848**, L16
- Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. 2013, *Natur*, **500**, 547
- Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. 2015, GCN Circ., 18100, <https://gcn.gsfc.nasa.gov/gcn3/18100.gcn3>
- Tanvir, N. R., Levan, A. J., González-Fernández, C., et al. 2017, *ApJL*, **848**, L27
- Troja, E., Piro, L., van Eerten, H., et al. 2017, *Natur*, **551**, 71
- Tunnicliffe, R. L., Levan, A. J., Tanvir, N. R., et al. 2014, *MNRAS*, **437**, 1495
- Wallner, A., Faestermann, T., Feige, J., et al. 2015, *NatCo*, **6**, 5956
- Woosley, S. E. 1993, *ApJ*, **405**, 273
- Zhang, B., Zhang, B.-B., Virgili, F. J., et al. 2009, *ApJ*, **703**, 1696

Gamma-Ray Bursts

Andrew Levan

Chapter 7

GRBs as Cosmological Probes

7.1 A Range of Cosmological Probes

Primarily due to their extreme luminosity, GRBs are a powerful probe of many cosmological processes. This immediately makes them attractive cosmic lighthouses, since their light can be seen at extreme distances, even with relatively modest detector technology. Given this, it is unsurprising that GRBs are among the most distant objects known in the universe despite their relative rarity.

There are three main probes that can be used to access the early universe (e.g., $z > 5\text{--}6$). These are normal galaxies, quasars, and GRBs. They each offer rather differing selection functions and have advantages and drawbacks in their use. A detailed description of the cosmological insight possible from all these probes is well beyond the scope of this text, but the advantages and challenges of selection via these different probes are described briefly below.

- **Galaxies.**

Galaxies are extremely numerous and relatively luminous. They dominate the number counts in all deep-field images and thus have great statistical power. The vast majority of all sources observed (and confirmed) beyond $z \sim 6$ are galaxies. They also provide a direct measurement of the starlight that is contributing to the universe at that time and thus can provide a handle on the universal star formation rate, as well as the evolution of cosmic structure through studies of their luminosity function, structural parameters, and clustering. Galaxies are readily detected in optical/infrared observations and under certain conditions can be seen in the mid and far-infrared (via reradiated dust emission), as well as in the radio. Spectroscopic studies of galaxies provide gas-phase metallicities through forbidden line indicators. It is a fair statement that the majority of our knowledge of the high-redshift universe derives from studies of galaxies at these early universal epochs.

However, the principal drawback of galaxies is that they are faint. At $z \sim 6$, an L^* galaxy¹ has an absolute magnitude of $M_{1600\text{Å}} \sim -19$, corresponding to an apparent magnitude of $I \sim 28$. This is beyond the limit of most ground-based telescopes and can only be imaged with the *HST*. It is therefore well beyond the point at which spectroscopic observations yield reasonable signal-to-noise across a broad wavelength range. Instead, spectroscopic observations at these redshifts must focus on the detection of emission lines, most notably from Lyman- α . Although new techniques to improve the prospects for observations have been successful, perhaps most notably the use of observations of galaxies behind lensing clusters to increase the reach, it remains far from straightforward to obtain the detailed physical conditions within high-redshift galaxies. Furthermore, since galaxies have a luminosity function that appears to steepen at high redshift, most of the light at $z > 6$ arises from galaxies that are too faint to be seen, even in our deepest images.

- **Quasars.**

Quasars are extremely bright and long-lived (probably hundreds of millions of years) objects powered by high accretion rates onto a supermassive black hole. They have a luminosity such that they can be observed at the highest redshifts with telescopes with moderate apertures. This in turn means that when using larger telescopes (e.g., 8 m class), their light can be dispersed to obtain detailed absorption measurements, detecting light across the optical and infrared window and providing high-quality information in absorption. As such, quasars have proved to be powerful tools in unveiling the properties of the intergalactic medium at high redshift, in particular via the detection of the so-called Gunn–Peterson trough, a signature of the end of cosmic reionization.

However, the use of quasars also poses challenges. First, they rely on the presence of massive black holes that should take some time to build up after the Big Bang. Indeed, the most luminous distant quasars contain black holes with masses in excess of $10^9 M_\odot$, something very challenging to create (Mortlock et al. 2011). However, it is clear that such quasars are rare, so it is difficult to locate them in large numbers, even with sensitive all-sky surveys.

Furthermore, quasars are highly luminous events that influence their environments because of their emission; hence, the lines of sight to quasars may not be indicative of the line of sight to a typical star, especially when investigating in proximity to the quasar itself. Finally, quasars tell us little about what stars are doing because they are so bright that they overwhelm the light from their host galaxies. To obtain any information about the stars with this host requires the complex subtraction of the quasar light and so far has not provided useful information in very high redshift examples.

¹ Here L^* is a characteristic luminosity that is defined as the break in the luminosity function. At luminosities lower than the break, the number counts follow a power law with a faint end slope ($L^{-\alpha}$ with $1 < \alpha < 2$ in most cases), and above L^* , the number counts fall exponentially.

- **GRBs.**

GRBs bypass several of the problems that exist in using galaxies and quasars as high-redshift probes. They are sufficiently bright, at least at peak, to be visible to the highest redshifts and for absorption spectroscopy to be plausible, sometimes for days after the burst (see Figure 7.1 for a comparison of the luminosity of GRBs, quasars, and galaxies). Because they are brief, they have a limited impact on the galaxy and environment around them. They are linked to the collapse of massive stars (Hjorth et al. 2003) and thus provide a census of star formation and star-forming galaxies at a given redshift. The burst afterglow also fades away, providing a direct view of the underlying galaxy, without the difficult task of subtracting the (much brighter) point source that exists when working with quasars. Finally, since the presence of the GRB provides evidence for at least one massive star at that position, GRBs can provide information on the potentially significant

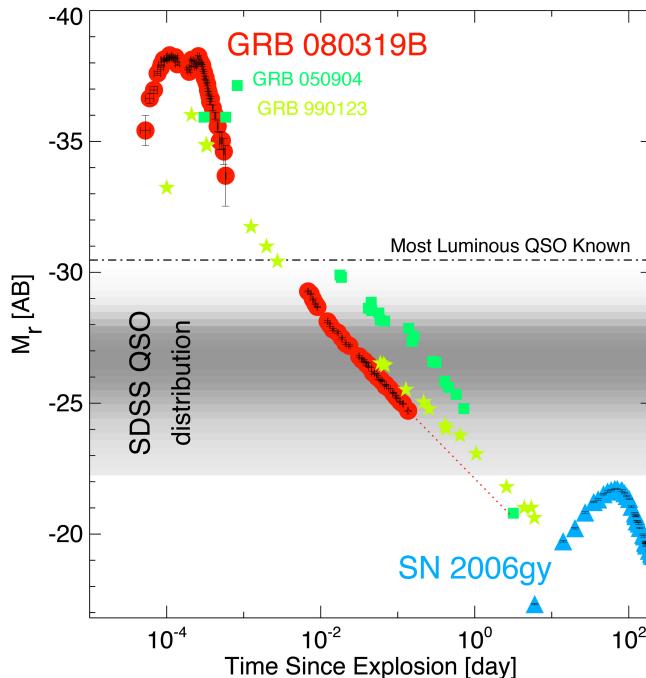


Figure 7.1. Comparison of the brightness of various cosmic probes from Bloom et al. (2009). Typical galaxies are not shown but lie at the lower end of the distribution, with absolute magnitudes of $M_r \sim -21$ at L^* . The SN 2006gy is a superluminous supernova and remains one of the most luminous supernovae ever seen. The distribution of magnitudes of the known quasars is shown in the shaded region, while some individual GRBs are marked. The difference in magnitude on the y-axis corresponds directly to the difference in apparent magnitude of the source at a given redshift, showing that GRBs can be a factor of 10 million brighter than a typical galaxy and 1000 brighter than a quasar. However, since they are fleeting explosions, the time to exploit this brightness difference is small, and by timescales of a day, most GRBs have magnitudes comparable to QSOs and are only as bright as galaxies on timescales of a few days. © 2009. The American Astronomical Society. All rights reserved.

fraction of galaxies at high redshift that are too faint for detection in the deepest observations currently available.

That is not to say that GRBs do not also have drawbacks. It is possible that there are complex selection effects that impact which stars form a GRB. It seems likely that this can be reduced to a simple metallicity dependence, but other effects may also be at play. GRBs at high redshift are also rare, and those beyond $z \sim 5$ still number only in the tens, compared to several hundred for galaxies. Finally, GRB science relies on rapid response while the afterglow is bright. Obtaining the level of information that is possible from quasars is therefore very challenging from a GRB afterglow.

7.2 Science from High- z GRB Afterglows

The most straightforward diagnostic from a GRB afterglow is the redshift of the burst itself. In many cases, it is possible to obtain high signal-to-noise observations that allow the detection of a wealth of spectral lines, both in the host galaxy and along the line of sight between the burst and us, providing robust redshifts and information about the physical conditions close to the burst. In other cases, when the afterglow is faint, as is often the case at high redshift, it is necessary to use bulk features to identify the likely redshift.

7.2.1 Selection of High- z GRBs

The most robust route to obtain a redshift for any object is to measure it directly via a spectrum. The identification of multiple emission or absorption lines leaves little doubt as to the identification of the burst redshift. Indeed, in most cases where good-quality spectra are obtained, the only route through which the burst redshift could be misidentified is if it actually lies behind the galaxy seen in absorption.

The selection of high- z GRBs largely follows one of the same approaches used very successfully for galaxies and quasars, namely the Lyman-break technique (Steidel et al. 1996). The idea of the Lyman-break technique is that pockets of neutral hydrogen along the line of sight will strongly absorb incident ultraviolet light. In particular, since most neutral hydrogen sits in the ground state, photons incident upon the hydrogen with wavelengths of 1216 Å (the wavelength of Lyman- α , the $2 \rightarrow 1$ transition in hydrogen) will be absorbed, exciting the H atom. Any photon with a wavelength < 912 Å (the ionization potential of hydrogen) will also be absorbed, ionizing the H atom. Since these hydrogen clouds are spaced throughout the universe along the line of sight between us and a distant source, this will create a suppression of the optical light below $1216 \text{ \AA} \times (1 + z)$. Each cloud will introduce an absorption feature at $1216 \text{ \AA} \times (1 + z)$, resulting in a multitude of absorption lines often called the Lyman- α forest. Below $912 \text{ \AA} \times (1 + z)$, any incident photon can ionize the H (i.e., not just photons with a particular wavelength), and the absorption is complete. The strength of the suppression around $1216 \text{ \AA} \times (1 + z)$ depends on the number of clouds along the line of sight and increases at higher redshift, where the universe is smaller and the path length a photon must traverse from source to

observer is longer. Thus, for objects at very high redshift ($z > 5$), there is an effectively total suppression of visible flux at $1216 \text{ \AA} \times (1 + z)$.

This property of the intergalactic medium creates a straightforward route of detecting high-redshift objects. They should be visible in red filters, $\lambda_{\text{obs}} > 1216 \text{ \AA} \times (1 + z)$, and then not visible to much deeper limits in a bluer filter, $\lambda_{\text{obs}} < 1216 \text{ \AA} \times (1 + z)$; this is often called the veto filter. It is this technique that has enabled the detection of the majority of very high redshift galaxies observed to date in deep fields, as well as of distant quasars and GRBs. An example of the detections of two GRBs at extreme redshift is shown in Figures 7.2 and 7.3.

Possible Lyman-break signatures can be tricky to interpret, especially for galaxies. In particular, high-redshift galaxies are frequently faint, such that the difference in brightness between the veto filter and the filters in which a detection is made is sometimes sufficiently small that photometric scatter can result in false positives. Furthermore, for galaxies, there are a range of underlying spectral features that must be considered. The galaxy may be dusty and thus have a rapidly falling flux toward the blue anyway, or it may be old and exhibit a strong so-called Balmer break at around 4000 \AA (rest frame) due to absorption features in the atmospheres of intermediate-to-old stars. Indeed, in extremum, it is even possible to misidentify a brown dwarf as a distant galaxy based on its colors. This means that many high- z candidate galaxies have been challenged, and some have been shown to be incorrect.

Although GRBs are much rarer than galaxies, they sidestep the majority of these concerns. They are sufficiently bright at peak that deep observations in a veto filter,

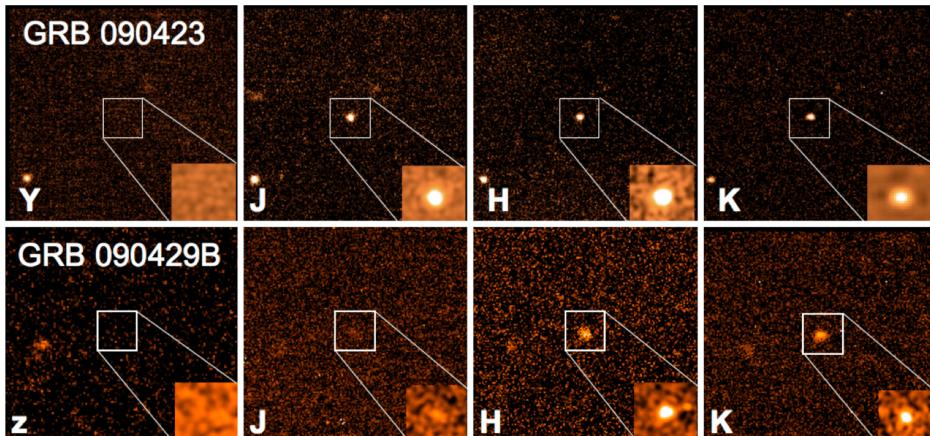


Figure 7.2. Discovery images of two very high redshift GRB afterglows in the z (850 nm), Y (1000 nm), J (1250 nm), H (1600 nm), and K (2200 nm) bands. The top panel shows the most distant spectroscopically confirmed burst, GRB 090423 (Tanvir et al. 2009), while the bottom panel is the highest photometric redshift, GRB 090429B (Cucchiara et al. 2011). The presence of the Lyman- α break can clearly be seen in both image sets. In the case of GRB 090423 ($z = 8.23$), it lies between the Y and J bands and is responsible for the strong drop in the afterglow brightness between these bands. In GRB 090429B, it lies in the J band and thus explains the faint detection of the counterpart in J (and its nondetection in z). Reprinted by permission from Springer Nature: Tanvir et al. (2009).

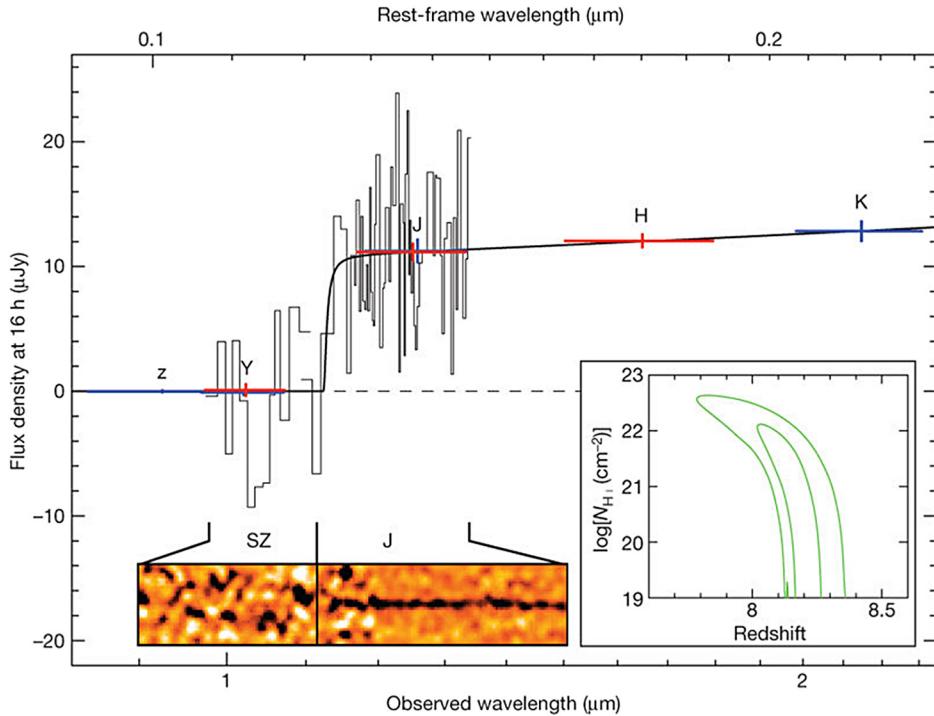


Figure 7.3. Spectroscopic and photometric observations of the afterglow of GRB 090423 at $z = 8.23$. The strong break at 1100 nm is clearly visible. The inset shows an image of the spectra (rather than its one-dimensional extraction), as well as the confidence intervals for the redshift, which depends weakly on the hydrogen column density along the line of sight to the burst. Reprinted by permission from Springer Nature: Tanvir et al. (2009).

combined with redder observations, can normally preclude dust reddening as the cause of the observed color. They are variable, so there is no doubt about the identification of the source, and the GRB itself exhibits a very simple power-law spectrum. Hence, photometric redshifts from GRBs are typically more robust than those for galaxies or quasars.

Ideally, once a photometric search has suggested a high redshift for a GRB, a spectrum can be obtained. Such a spectrum can confirm the redshift by the direct identification of the Lyman- α break, as well as other possible absorption lines. However, the challenge for observing GRBs at high redshift is their fading. A GRB typically fades as t^{-1} , so observations must be obtained quickly if the benefits of the GRB brightness are to be maximized.

7.2.2 Metallicities and Hydrogen Column Densities

Beyond $z \sim 2$, the Lyman- α line is redshifted into the optical window at $1216(1+z) = 3648 \text{ \AA}$ and becomes above the atmospheric cutoff. This means that the optical spectra of GRBs taken from the ground can derive the neutral hydrogen

column density, N_{HI} , based on the depth of the absorption around Lyman- α . A large fraction of GRB afterglows in this range are so-called damped Lyman- α systems (DLAs) exhibiting column densities in excess of $N_{\text{HI}} = 10^{22} \text{ cm}^{-2}$, indicative of a line of sight that is passing through the dense central regions of the host galaxy.²

In addition to seeing the Lyman- α absorption, the majority of GRB afterglow spectra at these redshifts also exhibit narrow absorption lines from heavier elements such as carbon, silicon, sulfur, and iron. The combination of the hydrogen column density with these lines enables the metallicity of the gas to be measured in absorption for these distant systems. This is a particularly valuable probe of the distant universe. In particular, for most distant galaxies, the galaxies themselves do not have sufficient brightness to derive the metallicity of the material from absorption lines as done in GRB spectra; thus, metallicities are normally derived for these galaxies using emission lines of hydrogen and oxygen (see Chapter 8). While this is valuable, in general, it provides only a relative oxygen abundance, which can be converted to other abundances assuming some pattern. In contrast, absorption lines in concert with the hydrogen abundance can measure, for example, the ratio of iron to hydrogen (so-called [Fe/H]), which is more commonly used as a distinguishing feature in stellar evolution models. Indeed, when multiple lines are observed, it is possible to obtain abundances for different element groups, and this in turn provides greater information about the degree of chemical enrichment within the host galaxy and the different nucleosynthesis paths that may have been followed to achieve that enrichment.

7.2.3 Line-of-sight Extinction

The GRB afterglows have simple power-law spectra that are relatively well understood. This means that in the absence of any absorption, optical multicolor imaging or spectroscopy should show the GRB afterglow to be a power law in frequency (or wavelength), so $F_\nu \propto \nu^{-\beta}$. In practice, absorption modifies this, both by introducing signals of gas-phase absorption, such as Lyman- α or metal lines, and by changing the bulk shape of the spectrum due to dust absorption. In principle, if the intrinsic spectrum is well known, then the measured spectrum provides a direct measurement of the extinction law. This makes GRBs a potentially excellent route to determining extinction laws across cosmic time. Such extinction laws may change due to metallicity but also due to the physical conditions in the dust, such as temperature, density, etc., that may impact the dust grain size distribution. Significant effort has gone into measuring the extinction laws, with the majority of examples being consistent with relatively modest extinction from low-metallicity dust. In particular, the strength of the silicate feature at 2175 Å is normally consistent with an extinction law comparable to that seen in the Small Magellanic Cloud, rather than the Milky Way (Schady et al. 2007). However, there are examples of different extinction laws, including some that have strong 2175 Å features (Elíasdóttir et al. 2009) and others

² These DLAs are similar to those seen along quasar lines of sight. However, since they typically probe direct lines of sight to star-forming regions in the host galaxy, rather than other galaxies along the line of sight to a quasar, the distribution of hydrogen column densities appears quite different, favoring high- N_{H} sight lines.

that appear very different from any laws observed along sight lines in the Milky Way (Fynbo et al. 2014).

There is also a population of bursts where this line-of-sight extinction appears more extreme. In these cases, it is possible for the optical afterglow to be completely obscured by dust within its host galaxy. This is often quantified by the absence of an optical afterglow, the presence of a large absorption component in soft X-rays (the so-called X-ray N_{H}), and a limit to the optical afterglow that lies below plausible explanations for the GRB fireball, namely that the spectral slope between the optical and the X-ray is $\beta_{OX} < 0.5$ (Jakobsson et al. 2004).

Since GRBs are related to massive star collapses, extinction measures made along GRB lines of sight probe the fraction of star formation at a given epoch that is obscured (assuming that other selection effects on the stars that create GRBs are independent of the dust extinction, which is not necessarily a good assumption). This provides an independent route to measure obscured star formation that is largely complementary to the direct measurements made of dust reradiation by submillimeter observatories that first identified a significant obscured component at moderate redshifts (Blain et al. 1999).

7.2.4 Constraints on Reionization

In addition, GRBs can provide unique constraints on reionization: the transition from a neutral intergalactic medium formed by recombination at the formation of the cosmic microwave background (CMB) and the largely ionized universe that we observe today. Observations of polarization signals from the CMB pinpoint the first free electrons at the start of reionization and suggest that this process begins around $z \sim 9$ (Planck Collaboration 2016). On the other hand, the presence of strong absorption blueward of Lyman- α , the so-called Gunn–Peterson trough, observed in quasar spectra implies a small neutral fraction around $z \sim 6$ and suggests that the process was broadly complete at this time (Fan et al. 2006).

However, the details of reionization remain an open question. First, it is not clear how the process proceeded in cosmic time and along different lines of sight. In principle, since structure formation depends on the degree of overdensity, some regions may begin to reionize their environments before others. When this ionization does begin, it may proceed rapidly or slowly. Since material that is ionized can subsequently recombine, it is even possible that reionization happened more than once, perhaps because of multiple generations of stars. Therefore, mapping the process of reionization requires determining the neutral fraction across cosmic time and the sky.

There is also the question of the sources that provide the supply of ionizing photons. The requirement here is that there are sufficient photons both to ionize the universe and to keep it ionized, producing ionizing radiation for long enough that the universe at large becomes sufficiently large (the spacing of electrons and protons sufficiently big) due to cosmic expansion that recombination becomes unlikely. There are multiple possible sources for these photons. It seems that accreting supermassive black holes in the form of quasars are too rare to provide the bulk of the ionizing photons. However, starlight can only provide the necessary budget if the

vast majority of photons originate from galaxies too faint for direct detection with current technology and if a reasonable fraction of the ionizing photons can escape their immediate environment to ionize the intergalactic medium.

The GRBs offer information on all of these areas. Their bright afterglows can provide a measure of the neutral fraction directly from the shape of the absorption around Lyman- α . Increasing neutral fractions yield increasing absorption to the red side of the line itself (see Figure 7.4). Although this signature must be disentangled from local Lyman- α absorption in the host galaxy, as often noted by the presence of a DLA, the different shapes can be decomposed with sufficient signal-to-noise.

The afterglow spectrum can also provide a direct measure of the fraction of light that escapes to ionize the intergalactic medium; since the afterglows are naturally very bright, a direct measurement can be made of the flux (and spectral shape)

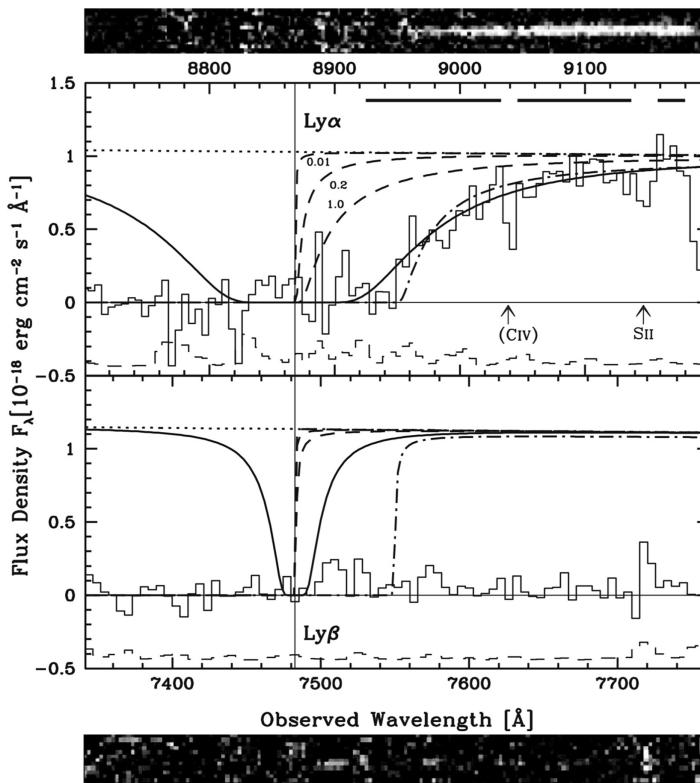


Figure 7.4. Optical afterglow spectrum of GRB 050904 (Kawai et al. 2006; Totani et al. 2006) taken 3.4 days after the burst. The cutout of flux around 8870 Å is due to the Lyman- α break, with the shape modified by the presence of a DLA and the state of the intergalactic medium. The dashed lines show how the shape of the break would vary in the absence of any DLA and suggest that high signal-to-noise observations can determine the ionization state of the universe at a given epoch direct from the afterglow spectrum. Reproduced from Totani et al. (2006). By permission of Oxford University Press on behalf of the Astronomical Society of Japan.

before the Lyman- α break. It is then possible to infer the expected intrinsic brightness at wavelengths shortward of Lyman- α and measure the fraction of this light that escapes directly from the spectrum. In most cases, only upper limits can be obtained, but these suggest that little light escapes in the vast majority of cases (Tanvir et al. 2018). This low escape fraction is challenging for reionization models that utilize entirely starlight.

Finally, GRBs can also provide a route to measuring the supply of ionizing photons. In particular, long GRBs are produced by the collapse of the massive stars that produce the bulk of the ionizing photon budget. The rate of GRBs thus provides some insight into the rate of photon production, although understanding the correction factors between the GRBs and the total star formation rate may be nontrivial. Further, GRBs select their host galaxies in a way that is, to first order, independent of the host galaxy brightness itself. This means that observations of GRB hosts can provide a direct means of measuring the galaxy luminosity function. Since we can place constraints even in the case where GRB hosts are not visible in deep images (because we know there was at least one massive star at that location), this provides a direct route to measuring the fraction of the ionizing photons that can be produced in galaxies too faint for inclusion in deep-field surveys and hence determining if starlight alone is sufficient to ionize the universe (Tanvir et al. 2012).

7.3 GRBs beyond $z \sim 5$

To the end of 2017, a total of \sim 12 GRBs have been located beyond $z = 5$, with 10 of these determined via spectroscopic redshifts and a further two via photometric techniques see Table 7.1.³

7.4 GRBs from Population III Stars

In principle, GRBs may be born from first-generation stars. These stars are expected to be born with much higher masses than the higher-metallicity Pop I and Pop II stars that are observed in the Milky Way, since cooling via molecular hydrogen in the pristine (pure H + He) environment of the early universe is much less efficient than cooling through line emission in heavier-element forbidden lines (e.g., O [II]) in a universe enriched by a previous generation of stars. At very high masses, these stars will evolve rapidly and can end their lives in a variety of ways depending on their composition, rotation, and initial mass. Many of these routes will not provide a visible GRB; for example, in very massive stars, it is likely that the temperature in the core will eventually exceed the threshold for pair production, yielding a so-called pair-instability supernova that completely destroys the star. However, it is plausible that in some cases, massive black holes can be formed in stars that have burned through the majority of their hydrogen via chemically homogeneous evolution (see Chapter 5). In this case, we might expect to observe GRBs, and these GRBs may be relatively bright (due to a larger mass reservoir in the more massive star) and of long

³These numbers are approximate. There are some marginal cases where the redshift is claimed to be high, but issues with signal-to-noise and foreground extinction complicate this interpretation.

Table 7.1. Sample of *Swift*-detected GRBs with Redshifts beyond $z = 5$

GRB	Redshift	Type
050814	5.4	Photometric
050904	6.29	Spectroscopic
060522	5.11	Spectroscopic
060927	5.5	Spectroscopic
080913	6.7	Spectroscopic
090423	8.2	Spectroscopic
090429B	9.4	Photometric
100905A	7.88	Photometric
111008A	4.99	Spectroscopic
120521C	6.0	Spectroscopic
130606A	5.91	Spectroscopic
140304A	5.28	Spectroscopic
140311	4.95	Spectroscopic
140515	6.32	Spectroscopic
160327	4.99	Spectroscopic

Note. Data taken from <http://www.mpe.mpg.de/~jcg/grbgen.html>.

duration (due to the larger size of the star). In principle, this might yield a distinct extra component of high- z GRBs, making them more common than might be expected based on pure extrapolations of the star formation rate (Bromm & Loeb 2006). To date, there is no clear distinction between the properties of high- z GRBs and those of low- z events.⁴ Hence, there is currently no evidence that we have observed GRBs from Pop III stars.

7.5 The Universal Star Formation Rate

The connection of long GRBs to massive stars suggests that they should be a route to determining the history of cosmic star formation. Indeed, if the factor relating GRB rate to star formation rate is a constant (i.e., a fixed fraction of massive stars create GRBs at all redshifts), then the volumetric GRB rate could be converted directly to the star formation rate. If the luminosity function of the GRB prompt emission is constant with redshift, then this becomes particularly straightforward, since one can simply determine the number of bursts above a given luminosity in a given redshift bin.

In practice, GRBs do not trace a fixed fraction of star formation and have a luminosity function of their prompt gamma-ray emission that may well not be constant with cosmic time. These two effects complicate but do not preclude the use of GRBs as star formation indicators. In particular, it is apparent that long GRBs have a metallicity bias and are only formed in galaxies (or at least environments within galaxies) with substantially subsolar metallicity. However, this can be well quantified

⁴Formally, the high- z events are more luminous, on average, but this is consistent with being purely a selection effect, where only the brightest systems are visible at high redshift.

at low redshift and, therefore, in principle removed, although doing so does require knowledge of the fraction of star formation contained within a given metallicity bin at a given cosmic epoch. More importantly, the threshold is such that at high redshift, where the largest uncertainties arise on the star formation rate from other methods, the universal metallicity is likely to have dropped to the point where the vast majority of, if not all, star formation lies below the critical metallicity for long-GRB production. Hence, beyond $z \sim 5$, there may be a direct conversion between the GRB rate and the star formation rate that does not vary with cosmic time. Indeed, similar statements also hold for the impact of the GRB luminosity function. Any changes that may occur, making GRBs brighter or fainter with cosmic time, are less likely to change rapidly at high redshift, since the GRB luminosity changes would presumably be related to the physical conditions in their progenitor stars. The only possible complication to this scenario would be a significant contribution from Pop III GRBs at high redshift, but since the properties of high- z GRBs are comparable to those of their lower-redshift cousins, this seems unlikely. Given this, it is unsurprising that attempts to use GRBs to measure the star formation rate produce excellent agreement with other methods at low redshift (Perley et al. 2016) and give confidence in their use at higher redshift.

Although there are systematic effects that must be considered when determining the star formation history from GRBs, when compared with other techniques, the uncertainties introduced by GRBs are generally of smaller order than those used by other techniques. For example, measuring star formation rates from Lyman-break galaxies requires two substantial corrections: the first is the correction for the fraction of star formation not contained within the observed sample, and the second is the correction for the presence of dust extinction. This thus requires the extrapolation of the star formation well below the detection limits of the survey. Since most high- z star formation takes place in these faint galaxies, this correction is very significant and could have implications if incorrect. Similarly, dust extinction is often an order-of-magnitude correction that must be applied to Lyman-break galaxy samples, as the UV light probed by Lyman-break searches is particularly sensitive to even small amounts of dust. Similarly, but in the opposite direction, attempts to derive star formation history from submillimeter galaxies must account for the galaxies that are not dusty as well. The issues that face GRBs when calibrating the star formation rate are therefore very different, so the agreement of the results from the different methods is encouraging.

7.6 Cosmological Parameters from GRBs

It is no surprise given the wide redshift range over which GRBs are visible that attempts have been made to utilize them as some form of distance indicator. Indeed, while Type Ia supernovae—the standard candle of choice—have yet to be found beyond $z \sim 2$, the presence of GRBs out to $z > 8$ means that differences in distances due to different cosmological parameters can be easier to distinguish, such that GRBs may be valuable standard candles even if the scatter in their intrinsic luminosities is relatively large.

At first sight, GRBs appear to be a very poor choice for a distance indicator; the range of apparent energies of the bursts spans six orders of magnitude, $10^{48} \text{ erg} < E_{\text{iso}} < 10^{54} \text{ erg}$. However, it is possible that relations can be obtained that shrink the scatter or that other observables can be used instead of the energy. Indeed, this may not be a hopeless scenario; Type Ia supernovae require standardization via a relationship between their luminosity and decay time before they can be used for precise cosmological work (Phillips 1993), and a similar relation for GRBs may be plausible.

The physical motivations for such a mechanism are much weaker than in Type Ia supernovae. In the supernova example, the explosion is somehow triggered when a white dwarf is pushed beyond the Chandrasekhar mass, either via a merger with another white dwarf or in accreting material from a main-sequence companion. This should be an approximately standard number. Since GRBs are massive star collapses, there is no reason to expect that massive stars undergoing collapse should be standard, and indeed, core-collapse supernovae in general are not. Despite this, much effort has gone into searching for such relationships, utilizing both the prompt emission and the afterglows at different wavelengths. Numerous possible relations have been uncovered, although the extent to which they represent true astrophysical relationships, selection effects, or chance probabilities is less clear. They broadly fall into two main categories, those that use the prompt gamma-ray properties and those that use the afterglow, and these are described briefly below.

7.6.1 Prompt Emission Relations

The first efforts to standardize GRBs originated from work on the prompt gamma-ray emission. Once the redshifts for bursts were known, it was possible to rapidly derive the inferred total energy or peak luminosity of the bursts in the bands observed by GRB-detecting missions. A slightly more complex procedure was then necessary to k -correct these values such that the bolometric energy and luminosity could be determined, ensuring that the properties measured were actually the physical properties of the burst and not due to either the restricted range of energy that the detector could observe or the shifting of the GRB spectrum due to cosmological expansion. Nonetheless, even before these more subtle effects were fully considered, it was clear that GRBs were not standard candles. The inferred range of isotropic equivalent energy⁵ spanned several orders of magnitude, with the majority of bursts lying in the region of $10^{52}\text{--}10^{54} \text{ erg}$ but with a significant tail extending down to 10^{48} erg . This meant that simply measuring the fluence or energy of a GRB was going to provide essentially no information on its distance.

Beaming Corrections

The first breakthrough in attempts to reduce this scatter to useful levels arose as observational evidence for beaming became stronger. Once the achromatic jet breaks

⁵ As noted in Chapter 1, this refers the energy the burst would have if it emitted the same energy we observe in all directions.

were seen in GRB afterglows and the opening angles of the jets could be estimated, it was possible to derive the true GRB energy emitted in gamma rays, since the total energy within the beam could, in principle, be estimated from the jet opening angle θ_j (e.g., $E_\gamma = f_b E_{\text{iso}}$, where the beaming fraction $f_b = 1 - \cos \theta_j \approx \theta_j^2/2$). Immediately, it became apparent that this energy was much more tightly correlated than the isotropic energies and had a strong peak around $E_{\text{true}} = 10^{51}$ erg (Frail et al. 2001). Although the scatter in this peak was still large, ~ 0.5 dex, there were still attempts to use the resulting energies for the determination of cosmological parameters. In part, this was motivated by the then relatively recent announcements surrounding the presence of dark energy as inferred from Type Ia supernovae. Clearly, an independent check would be valuable. However, although GRBs provided constraints that were consistent with those obtained from Type Ia supernovae, they were significantly inferior and thus could not meaningfully distinguish between different cosmological models (Bloom et al. 2003). Indeed, there remained doubts as to the extent to which GRBs could be treated as standard explosions.

Spectral Relations

Once it became clear that the apparent scatter in GRB energies could be reduced, further work was undertaken to ascertain if improvements could be made. In particular, investigations of the prompt emission began to search for correlations between observed parameters and the energy of the burst. Perhaps the best known of these relations is the correlation between the apparent peak of the νF_ν spectrum of the burst, E_{peak} , and the isotropic energy release, the so-called $E_p - E_{\text{iso}}$, or Amati relation (Amati et al. 2002; Figure 7.5). This relation broadly shows that more energetic GRBs emit a spectrum that peaks at higher energies and has the form (Amati 2006)

$$E_p = 95_{-9}^{+11} \left(\frac{E_{\text{iso}}}{10^{52} \text{ ergs}} \right)^{0.49 \pm 0.06}. \quad (7.1)$$

Although different analyses provide marginally different parameters, a similar relation was subsequently found between the peak energy and the beaming-corrected energy E_γ (Ghirlanda et al. 2004):

$$E_p = 267.0 \left(\frac{E_\gamma}{4.3 \times 10^{50} \text{ ergs}} \right)^{0.706 \pm 0.047}. \quad (7.2)$$

In addition to these relations, there are also relations based on timing, in particular, the relationship between the spectral lag (the time of arrival of different features in different high-energy observing bands) and the isotropic luminosity of the burst. For long GRBs, there is an apparent lag in the emission, such that the higher-energy emission arrives before the lower-energy emission, and it appears that this correlates with the luminosity of the burst. In principle, then, if one can measure the lag (and the redshift such that the lag can be placed on the rest frame of the GRB), the luminosity can be inferred following (Ukwatta et al. 2012)

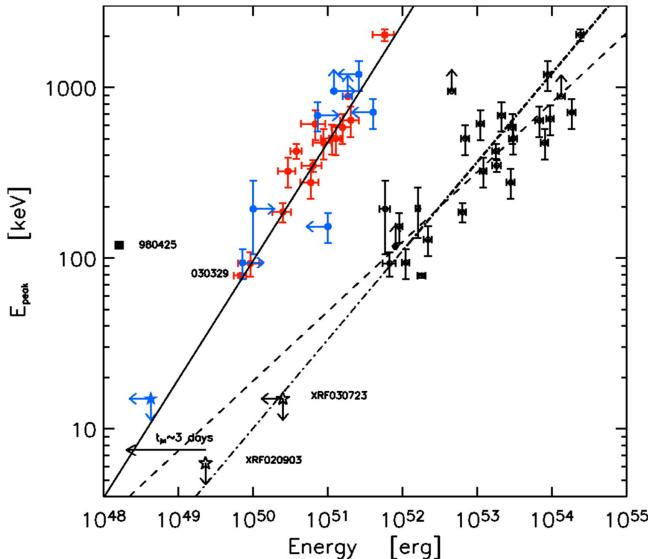


Figure 7.5. Amati, $E_p - E_{\text{iso}}$ (black), and Ghirlanda, ($E_p - E_{\gamma}$), relations from Ghirlanda et al. (2004). Both relations can be used to reduce the scatter in GRB energy releases and thus potentially as a route to standardizing the explosion. However, the Ghirlanda relation is substantially tighter and thus more promising in this regard. © 2004. The American Astronomical Society. All rights reserved. Printed in USA.

$$\log\left(\frac{L_{\text{iso}}}{\text{erg s}^{-1}}\right) = (57.4 \pm 0.4) - (1.2 \pm 0.2)\log\left(\frac{\text{Lag/(ms)}}{(1+z)}\right). \quad (7.3)$$

However, there was also considerable controversy over the extent to which these relations held for different populations of GRBs (long bursts, short bursts, under-luminous bursts, etc.) and if they represented true underlying relations in the GRB emission or could be explained (at least partly) by observational selection effects. Indeed, concerns were raised that such relations did not appear to be tighter when considering rest-frame properties, something that may be considered necessary for relations whole reflecting the underlying physics (Butler et al. 2009). Perhaps the strongest argument that these relations remain insufficient for cosmology is that to date, nobody has been able to derive a meaningful distance indicator for GRBs that can be used as a predictor of the redshift of the burst based purely on other observable quantities.

Afterglow Relations

With the success of prompt emission relations in narrowing the range of GRB energies, attention naturally turned to their afterglows. This provided a wide range of further parameters that could be investigated, albeit at the cost of unveiling statistical flukes or further selection effects.

Relatively early during the *Swift* mission, it was noted that the X-ray afterglows of many GRBs exhibited plateaus that lasted from a few hundred to several thousand s. Analysis of these plateaus suggested that longer-lived plateaus ended

at lower luminosities than shorter-lived plateaus (Dainotti et al. 2008). This leads to a relationship of the form

$$\log L_{X,p} = 48.54 - 0.74^{+0.2}_{-0.19} \log T_{X,p}, \quad (7.4)$$

where $L_{X,p}$ and $T_{X,p}$ are the luminosity and time at the end of the plateau, respectively (Dainotti & Del Vecchio 2017).

This relation has a significant scatter ($\sigma = 0.43$ dex) and thus is not useful for cosmological study. It lacks a clear physical motivation, although it is possible to recover the features with models where the late-time energy injection arises from a central magnetar (Rowlinson et al. 2014).

Indeed, there have been many other attempts to use this relation or variants thereof to improve inferences of GRB energetics, although they provide rather different parameters for the fits depending on the nature of the sample shown and do not suggest that they are likely to be of cosmological utility in the near future. A full summary is given in Dainotti & Del Vecchio (2017).

In addition to relations between the plateau luminosities and times, there have also been indications of optical relations between the prompt luminosity and the decay time. In particular, it appears that bursts with intrinsically more luminous afterglows normally decay more rapidly (Oates et al. 2012), providing yet another possible route to distance estimation. They found that

$$\log L_{O,200} = (28.08 \pm 0.13) - (3.636 \pm 0.004)\alpha_{O,>200\text{ s}}, \quad (7.5)$$

where $L_{O,200}$ and $\alpha_{O,>200\text{ s}}$ refer to the luminosity at 200 s postburst and the decay index, $t^{\alpha_{O,>200\text{ s}}}$, beyond this point. Again, this relation appears highly significant, but it may provide more insight into GRB jet physics than value as a route to the standardization of GRBs.

7.6.2 Afterglow–Prompt Relations

There are also a number of suggested relationships between the prompt emission and afterglow emission. Most fundamentally, there is a clear relationship between the observed GRB fluence and the brightness of the afterglow component (Gehrels et al. 2008; Nysewander et al. 2009), albeit with a very significant scatter. These relationships continue to hold when intrinsic properties are considered, and there are statistically significant relationships between the prompt E_{iso} and the afterglow luminosity in the X-ray (Berger 2007; Gehrels et al. 2008; Dainotti et al. 2011) and optical (Nysewander et al. 2009; Liang et al. 2010; Oates et al. 2015).

All of these relations, a summary of which is given in more detail in Dainotti & Del Vecchio (2017), have scatter that is much too large to use GRBs as useful direct routes to measure cosmological distances at the current time.

GRB Supernovae

Beyond the analysis of the properties of the prompt emission and afterglows, the variation in the properties of GRB supernovae has also been investigated. Early work in this direction suggested that the majority of GRB supernovae bore a strong

spectral resemblance to SN 1998bw and had a modest spread in their apparent peak magnitudes and rise times (Zeh et al. 2004). This picture has generally held up well as further examples have been found. Since all GRB supernovae appear to be of the same spectral type, there is at least some possibility that standardization may be possible. More recently, studies have looked for correlations similar to the famous Phillips relation that has been used so successfully to standardize Type Ia supernovae. In particular, this seeks to correlate the peak brightness of a supernova with its decay time. For Type Ia supernovae, the intrinsically brighter supernovae decay more slowly, and factoring this in allows the peak magnitude of the Type Ia supernova to be determined with dispersions of $\sim 10\%$ (Phillips 1993). The GRB supernovae appear to show a similar trend, although the number of good examples is far smaller than for Type Ia supernovae (Cano 2014; Li et al. 2014), so the scatter is yet to be well determined. In principle, the advantage that GRB supernovae have over Type Ia supernovae is that GRBs are readily detected at $z > 2$, where the detection of Type Ia supernovae has so far proved extremely challenging. The GRB supernovae are typically faint at these redshifts, with magnitudes comparable to Type Ia supernovae, and their spectral peak lies in the IR. Nonetheless, it is possible to measure GRB supernovae at this magnitude with space-based observatories such as *HST* and *JWST* (when it is launched).

7.7 The GRB Hubble Diagram

In summary, it seems likely that there are genuine correlations among GRB properties that suggest that the true underlying range of energies exhibited by GRBs is much smaller than indicated simply by looking at the apparent range of GRB energies. This should not be surprising. Indeed, it would be surprising if GRBs were produced in such a wide range of physical conditions that they could occur with energies spanning factors of 10,000 or more. Indeed, interpreting this variation as being due to physical processes, such as jet opening angles and Lorentz factors, structure within GRB jets, and viewing angles seems to be a very likely explanation for observing a broader diversity in energy than exists in nature. However, that GRB energies are more tightly clustered than inferred by the simplest observational diagnostics does not imply that there is an underlying system that is “standard.” Indeed, the physical processes most commonly invoked to explain GRB creation can clearly arise over a relatively wide set of conditions, and there is no strong explanation for why it should be possible to extract a standard energy. Given this, it remains highly uncertain that GRBs can be standardized to sufficient accuracy to be used either as a test or an independent check on cosmological parameters.

References

- Amati, L. 2006, *MNRAS*, 372, 233
- Amati, L., Frontera, F., Tavani, M., et al. 2002, *A&A*, 390, 81
- Berger, E. 2007, *ApJ*, 670, 1254
- Blain, A. W., Smail, I., Ivison, R. J., & Kneib, J.-P. 1999, *MNRAS*, 302, 632
- Bloom, J. S., Frail, D. A., & Kulkarni, S. R. 2003, *ApJ*, 594, 674

- Bloom, J. S., Perley, D. A., Li, W., et al. 2009, *ApJ*, **691**, 723
 Bromm, V., & Loeb, A. 2006, *ApJ*, **642**, 382
 Butler, N. R., Kocevski, D., & Bloom, J. S. 2009, *ApJ*, **694**, 76
 Cano, Z. 2014, *ApJ*, **794**, 121
 Cucchiara, A., Levan, A. J., Fox, D. B., et al. 2011, *ApJ*, **736**, 7
 Dainotti, M. G., Cardone, V. F., & Capozziello, S. 2008, *MNRAS*, **391**, L79
 Dainotti, M. G., & Del Vecchio, R. 2017, *NewAR*, **77**, 23
 Dainotti, M. G., Ostrowski, M., & Willingale, R. 2011, *MNRAS*, **418**, 2202
 Elíasdóttir, Á., Fynbo, J. P. U., Hjorth, J., et al. 2009, *ApJ*, **697**, 1725
 Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, **132**, 117
 Frail, D. A., Kulkarni, S. R., Sari, R., et al. 2001, *ApJL*, **562**, L55
 Fynbo, J. P. U., Krühler, T., Leighly, K., et al. 2014, *A&A*, **572**, A12
 Gehrels, N., Barthelmy, S. D., Burrows, D. N., et al. 2008, *ApJ*, **689**, 1161
 Ghirlanda, G., Ghisellini, G., & Lazzati, D. 2004, *ApJ*, **616**, 331
 Hjorth, J., Sollerman, J., Møller, P., et al. 2003, *Natur*, **423**, 847
 Jakobsson, P., Hjorth, J., Fynbo, J. P. U., et al. 2004, *ApJL*, **617**, L21
 Kawai, N., Kosugi, G., Aoki, K., et al. 2006, *Natur*, **440**, 184
 Li, X., Hjorth, J., & Wojtak, R. 2014, *ApJL*, **796**, L4
 Liang, E.-W., Yi, S.-X., Zhang, J., et al. 2010, *ApJ*, **725**, 2209
 Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, *Natur*, **474**, 616
 Nysewander, M., Fruchter, A. S., & Pe'er, A. 2009, *ApJ*, **701**, 824
 Oates, S. R., Page, M. J., De Pasquale, M., et al. 2012, *MNRAS*, **426**, L86
 Oates, S. R., Racusin, J. L., De Pasquale, M., et al. 2015, *MNRAS*, **453**, 4121
 Perley, D. A., Tanvir, N. R., Hjorth, J., et al. 2016, *ApJ*, **817**, 8
 Phillips, M. M., Aghanim, N., et al. 1993, *ApJ*, **413**, L105
 Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
 Rowlinson, A., Gompertz, B. P., Dainotti, M., et al. 2014, *MNRAS*, **443**, 1779
 Schady, P., Mason, K. O., Page, M. J., et al. 2007, *MNRAS*, **377**, 273
 Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, *ApJL*, **462**, L17
 Tanvir, N. R., Fox, D. B., Levan, A. J., et al. 2009, *Natur*, **461**, 1254
 Tanvir, N. R., Fynbo, J. P. U., de Ugarte Postigo, A., et al. 2018, arXiv:1805.07318
 Tanvir, N. R., Mackey, A. D., Ferguson, A. M. N., et al. 2012, *MNRAS*, **422**, 162
 Totani, T., Kawai, N., Kosugi, G., et al. 2006, *PASJ*, **58**, 485
 Ukwatta, T. N., Dhuga, K. S., Stamatikos, M., et al. 2012, *MNRAS*, **419**, 614
 Zeh, A., Klose, S., & Hartmann, D. H. 2004, *ApJ*, **609**, 952

Gamma-Ray Bursts

Andrew Levan

Chapter 8

Long-GRB Host Galaxies

Gamma-Ray bursts pinpoint the location of their progenitors and the host galaxies in which they reside. They are a unique way of identifying galaxies across cosmic history. Since they require only a single (or perhaps binary) star for their creation, the selection of the galaxy is not dependent on the brightness of the galaxy in some given observational band,¹ unlike the majority of routes to the selection of high-redshift galaxies. The GRB host galaxies can simultaneously provide direct constraints on the nature of GRB progenitors and unique insight into the nature and evolution of the galaxy population. In this chapter, the properties of GRB host galaxies are described, along with the physical conditions in which GRB progenitors must be formed and the cosmological knowledge that can be gained from studies of their hosts.

Stars form in galaxies; thus, when observing GRBs from stellar progenitors, it is natural that they should, by and large, arise from host galaxies. However, since the stars that create GRBs are not selected at random from all stars in the universe, the galaxies that host them are also not picked at random. In particular, since long GRBs arise from massive stars, we expect to observe them from galaxies in which such massive stars are present. Since massive stars exhaust their fuel reserves rapidly due to the high temperatures in their cores and the strong dependence of nuclear fusion rates with temperature, this means that we expect to observe long GRBs from host galaxies with young stellar populations. If there are other constraints on the nature of the progenitors beyond mass—for example, on their chemical makeup—then we also only expect to observe long GRBs from galaxies in which these conditions are met. Furthermore, if long GRBs occur through some even more exotic channel—for example, stellar mergers in dense clusters—we may expect to observe these in deep, high-resolution images of the host. Since the direct detection of GRB progenitors in pre-explosion images has yet to be performed (and likely

¹ This is not to say that GRBs select all galaxies equally, as will be discussed below, but that the detection of a GRB does not depend on the apparent galaxy brightness.

remains some years away at least), these diagnostics are perhaps the best route to inferring the necessary environmental conditions for GRB creation. Beyond this, these galaxies, which are located across cosmic history ($0.01 < z < 9.5$), provide a picture of the locations of the stars necessary to create GRBs across cosmic time. In the simplest model, where GRBs occur from some fixed fraction of massive stars, this means that they provide a snapshot of cosmic star formation that potentially is far less affected by observational biases than other approaches. Even if environmental effects alter this fraction with redshift, it is still possible to obtain novel constraints on the star formation history, especially if the nature of such biases can be directly measured in low-redshift samples of host galaxies. Finally, GRBs also shine through their host galaxies, revealing them in absorption, even when they are too faint to be detected in emission. In principle, this means we know more about many (sometimes invisible) GRB host galaxies than high-redshift galaxies selected via any other technique and that GRBs enable the construction of galaxy samples in which our knowledge of their physical properties is highly complete.

This chapter describes observations of GRB hosts, predominantly as observed by their emission, but also includes insights obtained in absorption via spectroscopy of their afterglows (see also Chapter 7). It outlines how the properties of the hosts themselves and the locations of bursts upon them provide unique routes to characterizing GRB progenitors (see also Chapter 5) and highlights what can be learned from multiwavelength observations. This chapter does not describe short-GRB hosts, which are explored in Chapter 6.

8.1 Early Observations

The first host galaxies were identified shortly after the first GRB afterglows. Perhaps remarkably with hindsight, these were not uniformly accepted as host galaxies on first sight, and there was debate about the nature of the apparent nebulosity underlying GRB 970228 as it faded (Figure 8.1; Sahu et al. 1997), with some suggestions that it may be related to a supernova remnant. Such suggestions gained credence by (now clearly incorrect) claims of apparent proper motion in the afterglow itself (Caraveo et al. 1997). Once redshift measurements became available with GRB 970508 (Metzger et al. 1997), it became apparent that the nebulosity must relate to an underlying galaxy, and indeed, a redshift was ultimately acquired directly for the host of GRB 970288, $z = 0.695$ (Bloom et al. 2001).

The first observations began to yield some information as to the nature of the GRB hosts: they were typically faint, often compact, and required deep observations to unravel in any detail. In addition to redshifts from the afterglows themselves, it was often possible to obtain redshifts from the host galaxies via observations of strong emission lines excited by ongoing star formation in the hosts (e.g., [O II], [O III], H α , [N II]). The galaxies were also typically blue and thus were clearly star-forming systems. Indeed, from the sample of GRB hosts built through the late 1990s and early 2000s, it became apparent that the host galaxies of all long GRBs appeared to be star-forming (e.g. Christensen et al. 2004). This was a situation strongly reminiscent of core-collapse supernovae, which appear exclusively in star-

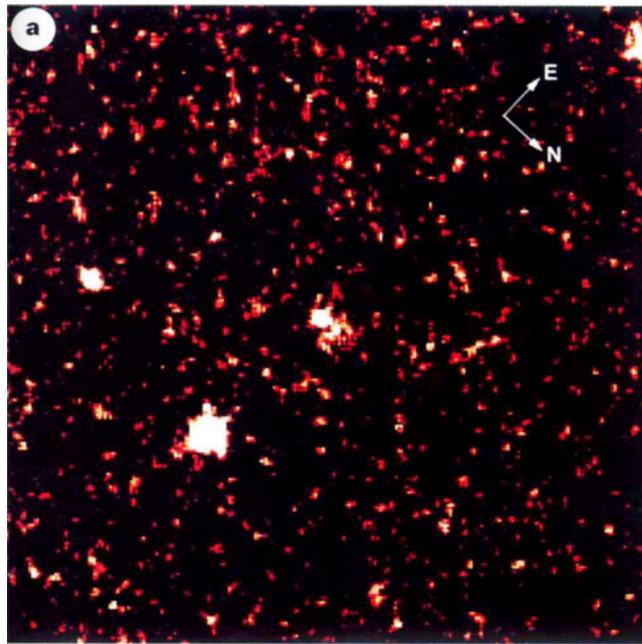


Figure 8.1. Observations from *HST* of the “first” GRB host galaxy, GRB 970228, from Sahu et al. (1997). The faint nebulosity noted from the ground is clearly visible and appears to be a faint host galaxy. Reprinted by permission from Springer Nature GmbH: Sahu et al. (1997).

forming hosts, while Type Ia supernovae (born from the explosions of white dwarfs) occur in both old and young populations (see Figure 8.2). This observation thus provided strong evidence of the link of long GRBs to massive stars before the supernova connection was firmly established.

8.2 GRB Hosts in the Galaxy Zoo

The standard route to the initial detection of distant galaxies is deep imaging. Traditionally, this has been done in the optical or IR, where the stellar populations of galaxies are at their brightest. Most often it is done via broadband imaging and the application of the Lyman-break technique (see Chapter 7) in order to ascertain approximate redshifts. However, surveys in narrow bands, in particular looking for the presence of Ly α , have also had success. The advent of new technologies at longer wavelengths meant that the 1990s also opened the window on the detection of galaxies, not via their stellar emission but via light reradiated via dust, thus creating an additional route of identifying so-called submillimeter or SCUBA galaxies.²

²The name SCUBA refers to the Submillimetre Common User Bolometer Array on the James Clark Maxwell Telescope, where many of the early detections of such galaxies were made. It is more common today to hear these galaxies referred to as submillimeter galaxies, given the more varied routes to their discovery that are now used.

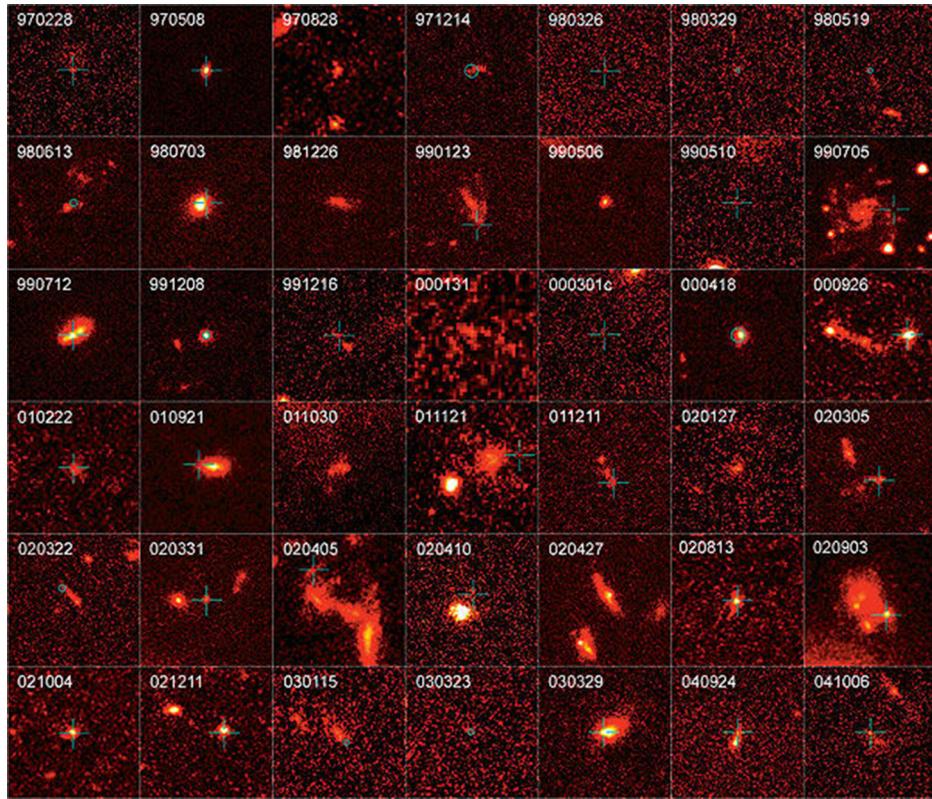


Figure 8.2. Imaging from *HST* of a large sample of long-duration GRB host galaxies from Fruchter et al. (2006). The hosts can be seen in a wide range of galaxies, although there is a notable paucity of grand-design spirals, with only the host of GRB 990705 appearing to arise from a spiral galaxy. In contrast, core-collapse supernova hosts at a similar range of redshifts exhibit a spiral fraction of almost 50%. These observations were the first to show strong differences between normal core-collapse supernovae and GRB environments. Reprinted by permission from Springer Nature: Fruchter et al. (2006).

All of these techniques rely on some intrinsic property of the galaxy for their detection. There must be sufficient luminosity from stars to reach some observational limit at a given redshift; for submillimeter galaxies, this must be combined with a sufficient mass of dust so that the reradiated emission can be detected. The samples of galaxies obtained from these surveys have proved extremely powerful but only contain a fraction of the true galaxy parameter space because of their observational constraints. For example, a galaxy can only be identified as a Lyman-break galaxy if it means stringent cuts on its apparent colors, in particular vanishing at wavelengths shortward of the rest-frame Ly α emission and then appearing blue at wavelengths longer than this emission. These restrictions are necessary to avoid the inclusion of contamination from red sources at different redshifts than those being probed (for example, a $z \sim 6$ galaxy might be confused

with a dusty or old galaxy at $z \sim 2$). However, this also means that any dusty galaxies at $z \sim 6$ are omitted from this survey, either because they do not meet the necessary color restrictions or because they are simply too extinguished to be detected in the first place. Conversely, the presence of large dust masses is a prerequisite for the identification of a galaxy via its submillimeter emission, since the heated dust is the source of the observational signal, and galaxies without dust cannot readily be detected. Building a complete picture of the galaxy population at a given redshift then requires stitching together the picture painted from these disparate observational samples and correcting for the selection effects imparted by each one. This is possible and has been very successful, but it is not the only route to understanding galaxies and their evolution at high redshift.

In principle, GRBs bypass many of these concerns because they depend only on a stellar-scale progenitor. Hence, they can provide information about their host galaxies, even when the host galaxies themselves are not detected (see Figure 8.3). For example, if most of the star formation at $z > 6$ is in galaxies beyond the reach of *HST*, then we would expect to observe the majority of GRBs at that redshift without apparent host galaxies (i.e., not detectable with the limits available), meaning that GRBs can provide information about galaxy properties across the luminosity function irrespective of the redshift of observation (Tanvir et al. 2010). This is particularly important at high redshifts, where observations with *HST*, and even in the near to mid-term future with the *James Webb Space Telescope* and the next generation of extremely large ground-based telescopes, will only probe the tip of the iceberg. Indeed, since the shape of the galaxy luminosity function evolves strongly with redshift (Bouwens et al. 2007), an increasingly large fraction of all stars are formed in small galaxies at progressively higher redshift, making these insights of particular value. Indeed, in principle, GRB spectra can provide great detail about these very faint, distant galaxies (see Figure 8.3).

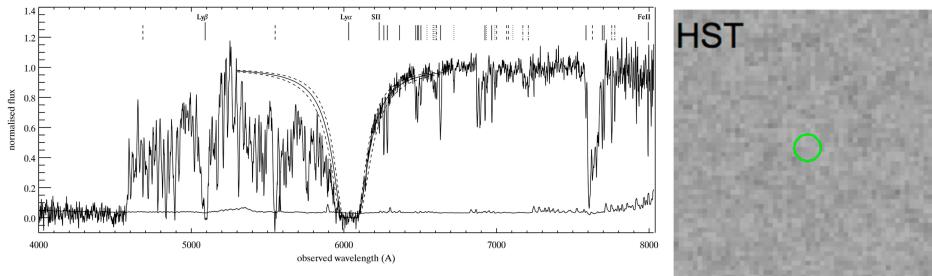


Figure 8.3. Afterglow spectroscopy and *HST* imaging of GRB 050730. The left-hand panel shows a high signal-to-noise afterglow spectrum (from Starling et al. 2005), demonstrating the presence of a damped Ly α system and several narrow absorption lines that enable the measurement of the GRB metallicity (which in this case is very low, $< Z_{\odot}/10$). The right-hand panel shows an *HST* image of the field in which there is no detection. In this case, the GRB afterglow has enabled the measurement of the redshift and physical properties of a galaxy that would not have been included in observations obtained via any other technique. Reproduced with permission from Starling et al. (2005) © ESO.

8.3 Basic Properties of Long-GRB Hosts

Although there are a wide range of properties observed for long-GRB host galaxies, there are a number of properties that can be regarded as somewhat prototypical of the class. In particular, we would expect a long-GRB host galaxy to be (Christensen et al. 2004; Fruchter et al. 2006; Savaglio et al. 2009; Perley et al. 2016a)

- star-forming, with a relatively high specific star formation rate (star formation rate per unit mass);
- compact or irregular, with a half-light radii of a few kpc or less (we observe very few grand-design spiral host galaxies); and
- low metallicity, especially at the site of the burst.

8.4 Building Meaningful Samples of GRB Hosts

Detection of GRBs is now routine. *Swift* continues to detect approximately two bursts per week and has done so for more than a decade. The result is a set of over 1000 GRBs with a roughly uniform detection mechanism scattered across the sky and throughout cosmic time. In principle, this creates the opportunity to perform detailed statistics on the properties of the host galaxy population, with sufficient numbers to be directly compared with samples of galaxies selected via other techniques (Lyman-break galaxies, Lyman- α emitters, submillimeter galaxies, etc).

However, such work is not without pitfalls. In particular, GRBs occur across the sky at random times. They are not located in surveys that optimize the detectability of the host galaxies, unlike deep pointed surveys that find galaxies via other techniques. For example, some bursts are lost because they lie close to the galactic plane, where foreground absorption and crowding preclude the identification of the host. Others are close to the Sun or Moon when discovered and limited, or no follow-up is obtained. Indeed, since the GRB fades and redshifts are often obtained only from the afterglow, even the weather at major observatories could, in principle, affect the information available for a given GRB. Hence, using all of the bursts detected by *Swift* would provide a significantly incomplete view of the GRB host population. Alternatively, selecting a subsample based on the presence of some additional information (for example, the burst redshift) introduces selection effects that are complex to understand and could have significant implications for the conclusions drawn. For example, in the case of requiring the burst redshift, a strong bias would exist against bursts for which there was no optical afterglow (the so-called dark bursts); this may lead to the incorrect conclusion that there is relatively little obscured star formation in the universe.

To avoid these biases, there have been various attempts to create objective criteria and define populations of GRBs that are prime for further study. In particular, one can first remove features that impact the observability of the GRB but are extrinsic to any properties of the GRB themselves; for example,

- GRBs with high foreground extinction,
- GRBs at extreme declinations where observatories cannot readily point,
- GRBs very close to the Sun or Moon when they occurred, and

- GRBs for which bad weather or technical issues precluded follow-up observations (but importantly, not those for which no follow-up was attempted because of a lack of interest).

This can be carefully coupled with additional requirements that do take into account the nature of the burst (at least to some degree), such as

- a limit on the peak flux or fluence of the burst, so that there is confidence that similar bursts would always be detected by the satellite, and
- the detection of an X-ray afterglow that localizes the burst to a few arcseconds.

Numerous samples have been built that have utilized some or all of these criteria. Doing so greatly enhances the observability and hence the redshift completeness (see Figure 8.4). The Optically Unbiased GRB Host survey (TOUGH; Hjorth et al. 2012) used a series of cuts to choose bursts with improved observability, creating a sample that now has $\sim 90\%$ redshift completeness (Jakobsson et al. 2012; Krühler et al. 2012).

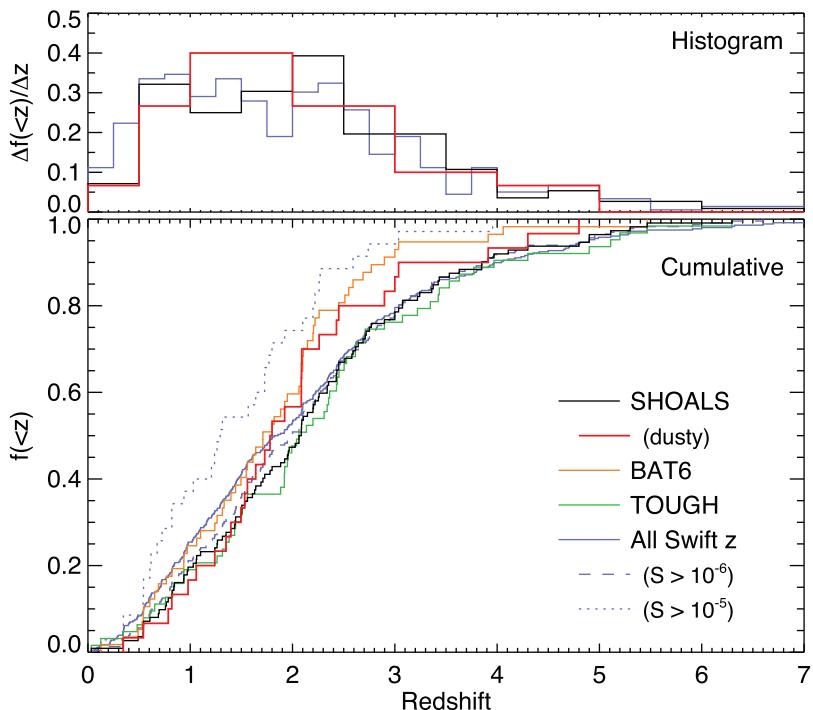


Figure 8.4. Redshift distribution of GRBs from Perley et al. (2016a). The redshifts are shown for all *Swift* bursts and then broken down into the various attempts that have been made to create “complete” samples of GRBs by restricting samples to those well placed for observation. There is an apparent bias to higher redshift that is introduced by this, although this may well be due to the increased redshift completeness of these samples. However, all redshift distributions broadly concur and find a median redshift of $z \sim 2$ for *Swift* GRBs. © 2016. The American Astronomical Society. All rights reserved.

The BAT6 sample looks at all GRBs with a fluence above 10^{-6} erg cm $^{-2}$. By looking only at the brightest GRBs, there is a far higher probability that they have X-ray afterglows, and indeed, this enables an impressive redshift completeness of $\sim 97\%$. More recently, these restrictions have been applied on an even larger sample of burst hosts as part of the *Swift* GRB Host Galaxy Legacy Survey (SHOALS; Perley et al. 2016a) project that now incorporates several hundred GRB host galaxies.

Between them, these surveys, which incorporate thousands of hours of observations with 8 m telescopes on the ground, as well as *Spitzer* and *HST*, mean that, for the first time, it is possible to draw robust conclusions about the global properties of GRB host galaxies across cosmic time and use them to investigate central remaining issues relating to the use of GRBs as cosmic probes.

8.5 GRBs Hosts at Optical and IR Wavelengths

Intensive observational efforts have been made to characterize the optical and IR properties of the host populations using the criteria described above. The TOUGH sample contains 69 bursts, BAT6 has 58, and SHOALS contains 119 (Perley et al. 2016a), with a new extended version containing >300 bursts. Each of these surveys adopted somewhat different approaches to follow-up observations, but the majority of the bursts have host galaxy observations obtained at 8 m ground-based telescopes (for example, the VLT for TOUGH) or the *Spitzer Space Telescope* (for SHOALS). Given the significant overlap in the samples, many of the burst have been observed as part of multiple surveys with different filters or survey design, providing excellent multiwavelength coverage with which to determine the properties of the galaxies.

8.5.1 Photometric Properties and Evolution

The majority of GRB hosts are detected in deep optical or infrared observations and have $R < 27$. Indeed, early observations of bursts (pre-*Swift*) with the *HST* enabled essentially all of them to be detected (Fruchter et al. 2006). Where available, these early observed galaxies show blue colors consistent with star formation (Christensen et al. 2004), and very few show significant evidence of dust extinction, for example, by appearing as so-called extremely red objects (Levan et al. 2006; Berger et al. 2007). The population of GRB hosts appears, on average, much less luminous than core-collapse supernova host galaxies (Fruchter et al. 2006; Svensson et al. 2010), which should, in principle, sample all star formation for stars with masses greater than the supernova-forming limit ($\sim 8 M_{\odot}$), consistent with their sampling a restricted fraction of the available galaxy population. However, other comprehensive analyses have argued that the host galaxy population was more typical (Savaglio et al. 2009), for example, with galaxies lying on the fundamental plane of stellar mass–star formation rate and metallicity (Mannucci et al. 2010; Campisi et al. 2011).

The majority of these early works were undertaken prior to the launch of *Swift*. At this era, GRB detection was sufficiently rare that it was impractical to cull samples based on, e.g., observability, and any well-localized burst was included for study. Indeed, the bursts themselves were detected by several different missions

(*BeppoSAX*, *HETE2*, *RXTE*, *IPN*, etc.) all with varying sensitivity, energy ranges, error-box sizes, and delays on source distribution. Hence, while some inferences from the population were clear and robust, others were less tractable. In particular, the paucity in the number of GRBs resulted in an inability to realistically study GRB host distributions, both within the distribution and across cosmic time. Doing so resulted in very few bursts per sample to consider and significantly weakened the results. For example, the work of Fruchter et al. (2006) compared GRBs to core-collapse supernova hosts at $z < 1.2$, a redshift bin spanning more than half the age of the universe. Even with the advent of much larger samples, much finer-grained observations remain challenging unless well designed.

The more recent studies, combined with advances in studies of the general galaxy population for comparison, provide a firmer and more compelling view. It appears that at low redshift, there is a clear paucity of GRBs in the most luminous galaxies (Perley et al. 2016c; Figure 8.5). This depletion exists even when the presence of dusty and extinguished GRBs is considered and is most likely related to a restriction on the metallicity of stars that can create GRBs (Vergani et al. 2015; Perley et al. 2016c). At higher redshifts, the inclusion of the dusty bursts enables GRBs to probe a much larger fraction of the available galaxy luminosity function and demonstrates that GRBs can be found in the more luminous hosts (Perley et al. 2013). There is an indication that the luminosity function of GRB hosts may undergo a change at around $z \sim 3$. For example, the TOUGH survey detects $\sim 80\%$ of host galaxies at $z < 3$ but only $\sim 25\%$ of bursts beyond this point (Hjorth et al. 2012; Schulze et al. 2015). If real, it is unclear if this distinction is due to the GRB progenitors or the

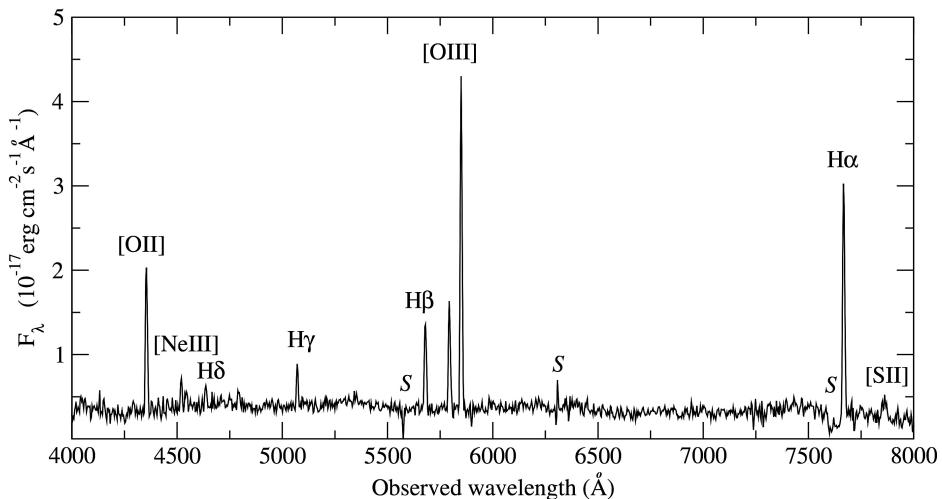


Figure 8.5. Optical spectrum of the host galaxy of GRB 030329 from Gorosabel et al. (2005). The spectrum is typical of those of GRB hosts and shows strong photoionized emission lines from hydrogen and forbidden-line emission from oxygen [O II] (3727 \text{\AA}) and [O III] (4959, 5007 \text{\AA}). Such luminous emission lines are common in GRB hosts and enable redshift measurements from host galaxies out to high- z . Reproduced with permission from Gorosabel et al. (2005) © ESO.

shape of the galaxy luminosity function. Indeed, the luminosity function is clearly changing rapidly at this epoch, and deep observations of several $z > 5$ GRB hosts with *HST* initially failed to uncover hosts, suggesting a steep faint-end slope (Tanvir et al. 2010) with the vast majority of star formation in galaxies beyond the limits of *HST*. While some high- z host galaxies have now been uncovered with instrumentation installed following the last *HST* servicing mission (McGuire et al. 2016), it does seem that the luminosity function at this redshift is indeed steep, such that there is little doubt that most of the star formation is in very faint galaxies.

8.5.2 Host Galaxy Spectroscopy

The optical and infrared spectra of GRB hosts are also highly characteristic. The majority of the spectra exhibit strong emission lines, including the Balmer series and forbidden-line emission from oxygen (see Figure 8.6). These lines have a high equivalent width and can be seen in emission even when the galaxy host is too faint for the spectrum to be well measured in absorption. This means that GRB host spectroscopy is plausible out to high redshift. Indeed, with current-generation technology, most notably the X-shooter spectrograph on the Very Large Telescope (that covers \sim 3000–25000 Å in a single shot), it has been possible to measure redshifts and even derive estimates of the metallicity of GRB host galaxies out to $z > 3$ (Krühler et al. 2015).

The high emission line fluxes are consistent with high star formation rates, as inferred from rest-frame UV imaging, and given the typically faint host galaxies, they imply a high specific star formation rate. The strong lines also enable metallicity measurements via various different indicators from strong lines, all of which point to a scenario where the majority of bursts arise from significantly subsolar metallicities when compared to the typical environments for star formation within the local universe (Graham and Fruchter 2013, 2017; see Section 8.7.1 for more details regarding the metallicity bias).

8.5.3 Morphology

The GRB hosts are typically compact, and aside from the most local examples, the majority are only marginally resolved at ground-based resolution. Insights into the host galaxy morphology therefore arise almost exclusively from observations with *HST*. These observations suggest that indeed, GRB host galaxies are extremely compact, with typical half-light radii of only \sim 1–3 kpc (Fruchter et al. 2006; Wainwright et al. 2007). Despite their faintness, they are also of high surface brightness (on average), consistent with a population with a high star formation rate in a starbursting mode.

In addition to visual searches, it is possible to use quantitative methods to define the morphologies of GRB host galaxies. A popular technique is the CAS system (Conselice 2003), which measures the galaxy concentration (the ratio of 20%–80% light radii ($C = 5 \log r_{20}/r_{80}$)), asymmetry (the size of residuals where the galaxy is rotated 180° about its centroid and subtracted from the original image), and clumpiness (determined by subtracting a smoothed version of the image from the

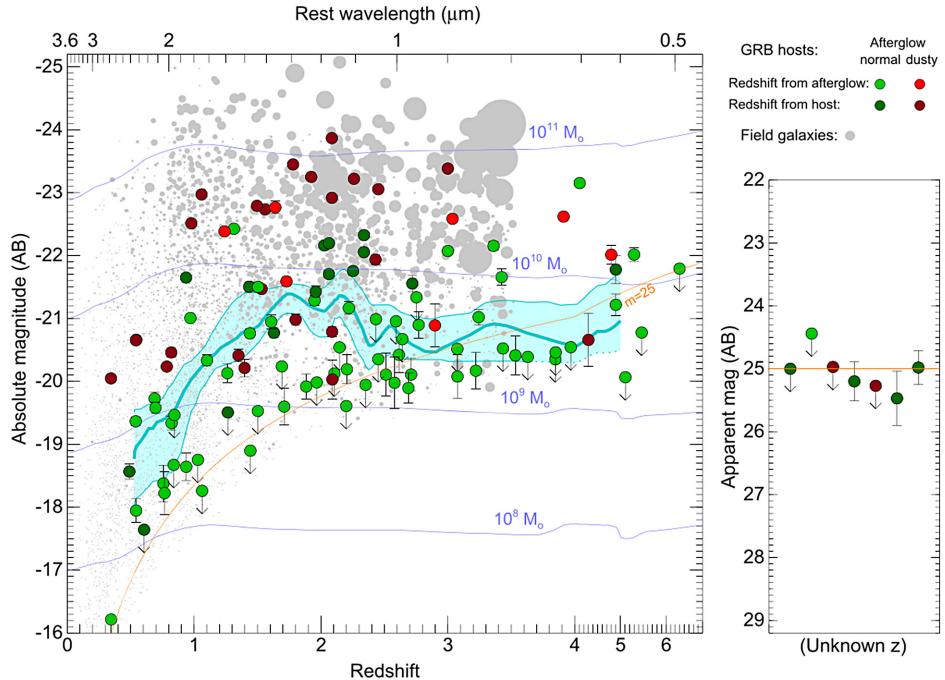


Figure 8.6. Redshift versus absolute magnitude diagram for GRB hosts and field galaxies (gray points) from Perley et al. (2016c). Red points are dust-obscured bursts, whereas green points are those that are (apparently) dust-free. Darker points (in each color) represent examples where the redshift measurement is from the host galaxy, and lighter shades are from the afterglow (note the prevalence of host redshifts for dusty galaxies). The background gray points are field galaxies, whose area represents their contribution to the star formation rate. Also shown is a fixed line corresponding to 25th magnitude in a given band. The paucity of field galaxies toward the lower reaches of this plot is therefore a consequence of observational selection effects, while GRB hosts in some cases have limits that may lie within this area. There is an apparent preference for GRB hosts to lie in lower-mass galaxies, and this is especially obvious at low redshifts. It can be modeled with a metallicity cutoff at around solar. The right panel shows the measured apparent magnitudes for the small number of hosts within the SHOALS sample that do not have measured redshifts. © 2016. The American Astronomical Society. All rights reserved.

original to identify high-frequency structure). Initial analyses of GRB hosts using these techniques suggested that they were spread across all of the galaxy types (elliptical, spiral, and mergers, based on the definitions for CAS originally proposed). In practice, the fraction ($\sim 30\%$) that were judged to lie in ellipticals were placed there because of a strong central concentration. They were clearly star-forming systems but were simply more compact than star-forming galaxies in the local universe today (Conselice et al. 2005). More recent work focusing on the IR (typically rest-frame optical) suggests that the majority of the host galaxies lie in the star-forming locus, with a modest number of mergers but few examples with both sufficiently large concentration and small asymmetry to be classed in the elliptical regime (Lyman et al. 2017).

8.6 GRB Hosts at Submillimeter and Radio Wavelengths

While the majority of studies of GRB hosts have focused on their properties at optical and IR wavelengths, there has also been intensive effort to understand their hosts at longer wavelengths, in particular in the submillimeter and radio. These regimes provide a different set of diagnostics to those available only in the optical waveband. In particular, submillimeter observations directly probe emission from dust within the host galaxies, while radio observations probe synchrotron emission from supernova remnants and provide a measurement of star formation activity that is independent of the dust extinction within the hosts. Therefore, the combination of these two regimes can test the extent to which cosmic dust impacts GRBs in terms of both the detectability of their afterglows (e.g., what fraction of afterglows are located in dusty hosts) and what fraction of GRBs (and hence potentially star formation) are highly obscured. This latter point is of particular importance, since corrections to the star formation rate to account for the impact of dust extinction or galaxy number counts are significant, and because deep observations of submillimeter galaxies imply that a significant fraction of all star formation at $z \sim 2$ is obscured.

Perhaps surprisingly, rather few GRB hosts are detected at longer wavelengths. This was particularly true during the period of early observations, and indeed, of the handful of claimed detections during this era, the majority appear to be due to late-time contamination from the radio afterglow or the chance alignment of unrelated galaxies in the (large) beam of submillimeter bolometers (Perley et al. 2017). This situation has improved recently with the upgrades performed to increase the bandwidth of the major radio observatories. These upgrades enable a much larger frequency/wavelength range to be covered and thus enhance the likelihood of detection, even when using the same dishes. It is an approach that has been very successful at both the Jansky Very Large Array (JVLA) and the Australia Telescope Compact Array (ATCA). These upgrades, combined with a large sample of GRBs, mean that the number of GRB hosts with secure radio detections now numbers in the tens. As might be expected, these detections do not appear to occur independently of the burst or optical properties of the hosts but are preferentially found in bursts that exhibited signs of dust extinction, such as a red or absent afterglow or red colors within the host galaxy. In part, this bias is explained by observational selection effects, in that these galaxies are those that have been pursued most extensively for observations. However, observations of more complete samples generally support the picture that detectable radio emission is the exception rather than the rule for GRB hosts (Michałowski et al. 2012; Perley et al. 2015).

Most recently, the capabilities of the Atacama Large Millimeter Array (ALMA) have also been brought to bear in this area. These observations provide far higher sensitivity than previously possible with submillimeter observations, thanks to a combination of increased collecting area and higher resolution available via interferometry. This capability has enabled stringent limits (a few $M_{\odot} \text{ yr}^{-1}$) to be placed on dust-obscured star formation in the most distant GRB host galaxies (Berger et al. 2014), as well as the more routine detection of molecular lines

(e.g., carbon monoxide (CO)) in host galaxies (Hatsukade et al. 2014), something subsequently confirmed through observations of other hosts at ALMA and elsewhere (Stanway et al. 2015; Michałowski et al. 2016; Arabsalmani et al. 2018). These observations provide a route to measuring the availability of molecular gas for star formation and comparing this with the dust at the same locations. The implication is that at least some GRBs are born in regions where the gas has been depleted significantly already, either because it has already been incorporated into stars or because the massive stars formed coeval with the GRB progenitor have dissipated it before the burst (Hatsukade et al. 2014).

8.7 GRB Hosts as Tools to Probe Progenitors

8.7.1 Metallicity Bias

A striking feature of GRB host galaxies noted early in their study was a preference for low-luminosity star-forming galaxies. Of particular note, a comparison of the host galaxies of GRBs and core-collapse supernovae at comparable redshifts (at this point in time, predominantly $z < 1$) showed that approximately half of the supernovae were occurring in massive, grand-design spiral galaxies; in contrast, only one (of approximately 40 known at the time) GRB host was a spiral (Fruchter et al. 2006). Indeed, the luminosities of the core-collapse supernova hosts were systematically higher, also implying higher stellar masses. There is a well-known correlation between the mass of a galaxy and its metallicity, perhaps well seen locally in the trends between the Milky Way, the Large Magellanic Cloud, and the Small Magellanic Cloud but in practice robustly established as a mass–metallicity (or luminosity–metallicity) relationship based on studies of many thousands of galaxies (Tremonti et al. 2004). Given this relation, the low masses of GRB hosts would naturally be explained by the presence of some kind of metallicity threshold for the production of GRBs, consistent with the expectations of the collapsar model.

Further studies of individual host galaxies added to this picture, showing that both the hosts at large and the locations of bursts within them appeared to originate at substantially subsolar metallicities. Gas-phase metallicities can be computed in such cases by measuring the flux (or equivalent width) of strong emission lines and comparing these to the Balmer series in order to obtain estimates of the metallicity. One such common approach used heavily in early observations of GRB hosts is the R_{23} measurement, which defines a metallicity via

$$R_{23} = \frac{F([OII](3727 \text{ \AA}) + F([OIII](4959 \text{ \AA})) + F([OIII](5007 \text{ \AA}))}{F(H\beta(4861 \text{ \AA}))}, \quad (8.1)$$

although this can be double valued and ideally requires additional diagnostics to break the degeneracies within (although these were often not available in the early days of GRB host studies). This early work demonstrated that many GRB hosts have metallicities of $Z = 12 + \log O/H < 8.3$ for the majority of these systems, compared to a solar composition of $Z_\odot = 12 + \log O/H = 8.8$; since this is a logarithmic scale, it implies that these host galaxies are typically drawn from a

population with metallicity $<0.3 Z_{\odot}$. Subsequently, more sophisticated analyses using updated diagnostics have been used, although these do not change the overall picture.

This conclusion stood up well to further observations taken over the course of the next decade or more. In most cases where a metallicity can be directly measured spectroscopically, it appeared that the metallicity was substantially less than the solar value, leading to the suggestion of a robust cut in metallicity around $1/3 Z_{\odot}$, perhaps with a relatively tight exponential cutoff. While this has become the consensus view, this picture was not universally acknowledged, and some authors noted that the galaxies lay well on a fundamental plane that incorporated stellar mass, star formation rate, and metallicity (Mannucci et al. 2010). These were highly star-forming galaxies, but such highly star-forming galaxies at the redshifts where GRB hosts can be seen tend to also have low metallicity (this can be thought of as a statement of so-called cosmic downsizing, where in the local universe, the majority of massive galaxies form few stars, and most star formation is in low-mass systems), and so it is possible to envisage other routes (beyond a metallicity bias) that may result in their formation in such locations, including unusually massive stars (that only form in very massive star-forming regions) or dynamical interactions.

However, this scenario was still incomplete because, in most cases, it omitted galaxies that had hosted dust-obscured bursts because these were harder to identify via their afterglows and were often faint at optical wavelengths, perhaps because of the effects of dust, which could also obscure their strong emission lines. Indeed, significant efforts to obtain precise positions and integrate these dusty galaxies into the global host population demonstrated that the hosts of obscured (dark) GRBs were in fact systematically dustier and more massive than those of the unobscured population (Perley et al. 2013), potentially casting doubt on the strength of the metallicity bias. Most in-depth observations enabled as part of the most recent SHOALS survey suggest (albeit using photometric masses, rather than directly measured metallicities) that there is indeed a strong threshold for GRB production, but that it should be drawn at the solar level, rather than $1/3 Z_{\odot}$ (Perley et al. 2016c). Indeed, large (but somewhat less homogeneous) spectroscopic surveys with the X-shooter instrument that covers the entire optical and IR window point to a broadly similar conclusion (Krühler et al. 2015). Hence, it seems very likely that metallicity does play a crucial role in the creation of a GRB, but the precise threshold for this effect remains unclear.

8.8 GRB Hosts as Tools to Probe Distant Galaxies

8.8.1 Ly α Emission

In addition to strong optical emission lines, emission from Ly α can also be observed in GRB hosts. This line enters the optical window at around $z \sim 2$ (i.e., when $1216 \text{ \AA} \times (1 + z) > 3000 \text{ \AA}$) and can, in principle, be observed in the optical to $z \sim 6$ and beyond in the IR. Early observations suggested that GRB hosts may show nearly ubiquitous Ly α emission, perhaps because of low metallicity and little dust within the host galaxies (Fynbo et al. 2003). Further observational efforts, in

particular those focused around studies of the more complete samples described above, fail to reproduce this result, with only $\sim 30\%$ of host galaxies exhibiting the line, which is more consistent with the expectations of typical star-forming galaxies at these redshifts.

8.8.2 The Host Galaxies of Dark and Dusty GRBs

The discovery of a population of GRBs whose afterglow light was apparently suppressed relative to the extrapolation of their X-ray afterglows (Groot et al. 1998; Rol et al. 2005) was met with a flurry of suggestions for their origin, including unusual processes in GRB jets, low ambient density media, very high redshift, and extinction by dust in the host galaxy (Perley et al. 2013). While all of these scenarios seem likely to be present in the observed samples to some degree, it now appears that in the majority of cases, the origin of the optical suppression is due to dust extinction within the GRB host (see Chapters 2 and 7 for details on the information obtained directly from the afterglows). Indeed, in addition to the dark bursts, there also exists a population of red afterglows from which the extinction can be measured directly and suggests in some cases that several magnitudes of optical extinction are present (Rol et al. 2005; Schady et al. 2007).

Such dusty lines of sight offer the ability to constrain obscured star formation. The presence of luminous submillimeter galaxies implies that the majority of star formation at $z \sim 2$ should be taking place in an obscured mode, so that optical and UV galaxy surveys fail to identify the star formation sites. The fraction of dark GRBs is somewhat poorly constrained, since limits are often not sufficiently deep to be meaningful, but it seems that a significant fraction (though $< 50\%$) of the bursts observed by *Swift* are indeed dark, which would be in keeping with a high fraction of obscured star formation (Cenko et al. 2009).

The ensemble properties of the host galaxies of these dark bursts are in keeping with this picture. While they span a broad range of luminosities, dark GRB hosts are systematically more luminous, larger, and redder than the host galaxies of the optically bright population. This is fully in keeping with the expectation of a dusty population of galaxies. Indeed, the GRB hosts that appear to be extremely red objects are almost exclusively from bursts with evidence for dust extinction. The inclusion of these galaxies in the overall samples of GRBs suggests that the population of GRB hosts does extend to higher luminosities (and hence potentially metallicities) than inferred from the optically bright hosts alone. However, even at $z \sim 2$, it is likely that the most luminous submillimeter galaxies are at a metallicity threshold beyond that which can sustain GRBs. Locally, there is indeed a paucity of IR bright host galaxies and dark GRBs, which may well be related to the locations of star formation at low redshift (Stanway et al. 2015; Perley et al. 2016a).

8.8.3 GRBs and the Star Formation Rate

Long GRBs are related to the deaths of massive stars. Since these stars live short lives (typically a few million yr), they explode as GRBs or supernovae in stellar nurseries in which stars are often still forming and the majority of stars are just at the

beginning of their lives (e.g., an average star has a mass less than the Sun and a lifetime greater than the age of the universe and thus is still very young at only a few million yr of age). This means that GRBs can directly probe the star formation rate. In the idealized assumption that all GRBs were seen and redshifts obtained, it would be trivial to go from a volume-averaged GRB rate to a star formation rate. In practice, this is not the case, in the local Universe much star formation likely takes place above the critical threshold for GRB creation. At earlier cosmic epochs the Universal metallicity evolution results in a larger fraction of star formation below this threshold. At very high redshifts, all star formation may be at a metallicity where GRB formation is allowed, and hence GRBs may become much closer to perfect tracers of star formation. Once the global metallicity of the universe falls below this threshold, the conversion rate between star formation rate and GRB rate should be fixed (under some other assumptions to do with initial mass function constancy), so the GRB rate would then provide a direct route to measuring the star formation rate. See Chapter 7 for further information.

8.9 Burst Locations and Environments

It is not only the bulk properties of the GRB hosts that provide valuable information on the nature of the progenitors and the role of GRBs as cosmological probes. The locations of the bursts on their hosts and the properties of the galaxies also carry much valuable information. Indeed, if sufficient resolution and signal could be obtained, then studying the stellar population directly under the burst position in principle provides more diagnostics than the host galaxy bulk properties.

Such work is not straightforward, however. At a typical GRB redshift of $z \sim 2$, the host galaxies are faint (integrated magnitudes of $R > 25$ are not uncommon), and the angular diameter distance is such that 1" on the sky is approximately 8 kpc. This means that at ground-based resolution, the scales within the host that can be resolved are comparable to the separation of the Earth from the galactic center.

Contributions from *HST* have been critical in unveiling the detailed properties of GRB locations, since its diffraction-limited performance yields a typical resolution of 0.1", a factor of 10 better than often possible from the ground. This enables burst host galaxies to be differentiated into multiple resolution elements, such that some information about the galaxy morphology and host location can be recovered. However, the faintness of the galaxies normally limits this spatially resolved information to broadband imaging. Only in the local bursts (typically $z < 0.2$) is there often sufficient spatial extent to conduct spatially resolved spectroscopy, and the new generation of wide-field integral-field spectrographs that yield spectra for each pixel (so-called spaxels) within a galaxy have made important inroads in this regard.

8.9.1 Quantifying Burst Locations

Perhaps the most straightforward measurement of a burst (or indeed any transient) location is to measure the offset of the transient from the center of its host galaxy.

This can be done readily, even from ground-based images, provided that an image exists when the GRB afterglow is bright that can be used to register the position of the afterglow on an image taken long after the afterglow has faded. These offsets alone provide some information about the nature of the progenitors, since young massive stars should explode where they are formed, while older stars, or especially those that have been kicked (such as neutron stars), should lie at much larger radii. It is common to quantify an offset relative to the size of the host galaxy itself. For example, in the Milky Way, stars extend to >10 kpc from the center, but this is not observed in the Magellanic Clouds. Therefore, a more valuable diagnostic is the host normalized offset (offset/host half-light radius). Major efforts were used to obtain this diagnostic for a sample of bursts in the years after the first afterglow discovery. Indeed, Bloom et al. (2002) demonstrated that approximately half of the GRBs arose from within the half-light radii of their host galaxies, indicative of a population that traced the stellar population (i.e., the light) within the host galaxies.

However, while this diagnostic is powerful, it is not always ideal. High-redshift galaxies are often irregular in morphology and contain multiple bright knots of emission. This means that the center of the galaxy can be complex to define and may on occasion have little physical meaning. This led to the desire to develop location diagnostics that were independent of the galaxy morphology, at least to first order. Such an approach was developed for GRBs using broadband *HST* imaging (Fruchter et al. 2006) and a similar method for narrowband images of supernovae (James & Anderson 2006). The principle of this technique, often referred to as the fractional flux, or F_{light} , is to sort the host galaxy pixels into rank order from faintest to brightest and then locate the pixel in which the transient occurred (in an image where the transient is absent) in the cumulative list of pixels. A value of $F_{\text{light}} = 0$ means that the transient does not lie on the light of its host galaxy, $F_{\text{light}} = 1.0$ is lying on the brightest pixel, and $F_{\text{light}} = 0.6$ means that 60% of the light of the galaxy is in pixels of surface brightness fainter than that in which the transient occurred.

The distributions of F_{light} for GRBs and core collapse supernovae are shown in Figure 8.7. It can clearly be seen that GRBs are highly concentrated on their host galaxy light, while supernovae broadly trace the light within their host galaxies. This means that GRBs preferentially occur on the very brightest pixels within the host galaxy, while the probability of a supernova in a given pixel is proportional to the surface brightness in this pixel.

The stellar populations within these distant host galaxies are not resolved into individual stars. However, by utilizing the expected distribution of stars within galaxies (spatially, within clusters, and in mass and luminosity), it is possible to estimate the masses of the progenitor stars that form GRBs. As described in Chapter 5, the luminosity of a star scales roughly as $L \propto M^3$, whereas the lifetime of a star scales as its mass/luminosity (since its luminosity is created by nuclear burning, some fraction of its mass) so that $t \sim M^{-2}$. In other words, the regions of a galaxy that host its most massive but shortest-lived stars are expected to be the brightest. One can then compare the distributions of F_{light} seen in GRBs and supernovae with those expected for stars of differing masses, leading to the conclusion that GRBs may arise from stars with masses $>40 M_\odot$ (Larsson et al. 2007; Raskin et al. 2008).

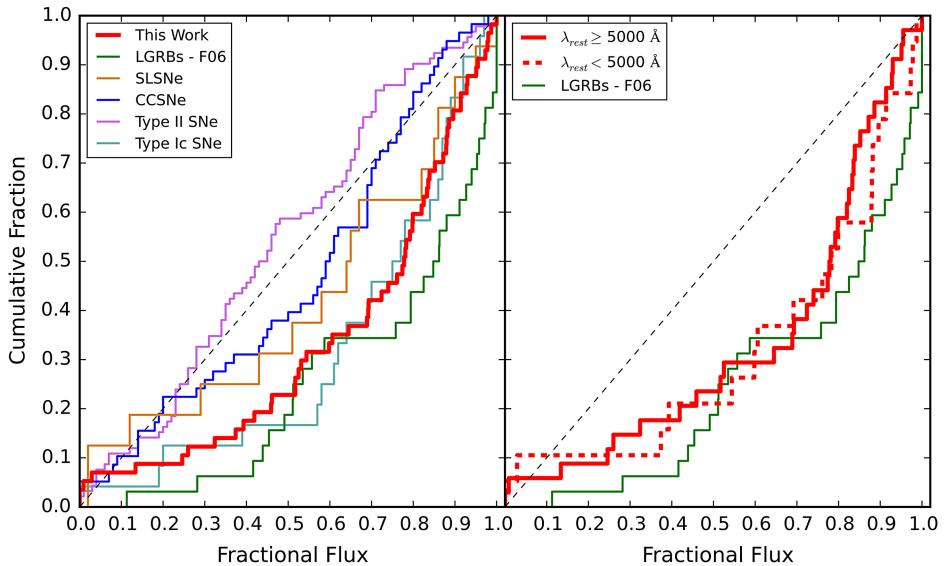


Figure 8.7. Fractional flux parameter (Fruchter et al. 2006) for GRB hosts from Blanchard et al. (2017). The left panel is for all GRBs, while the right panel distinguishes between bursts where the observed filter is either redward or blueward of 5000 Å. For blue light, it would be expected that massive star progenitors may linearly trace the light, at least to first order, and this is seen for core-collapse events (blue). In contrast, both Type Ic supernovae and GRBs appear to be more concentrated on the light, favoring the brightest regions. This is most likely because of a preference for their origin only in the most massive stars (Larsson et al. 2007). © 2016. The American Astronomical Society. All rights reserved.

8.9.2 Spatially Resolved Spectroscopy

While it is only available with current technology for relatively local galaxies, spatially resolved spectroscopy of GRB hosts has also provided valuable direct insight into the nature of the hosts and GRB progenitors. Most (but not all) of this work resolves around studies of the low-luminosity GRB population that are probably rather more common but much fainter than their more luminous cousins. This combination limits the horizon over which such bursts can be seen, so examples of low-luminosity bursts are almost exclusively found at low redshift, where the galaxies are relatively large on the sky. These galaxies are open to spatially resolved work, either by placing slits across the galaxy, perhaps at different angles, or more recently via the use of integral-field spectrographs, such as the MUSE spectrograph on the Very Large Telescope. These observations enable the measurements of stellar parameters across the galaxy. Importantly, they test whether gradients in star formation rate or metallicity seen in local GRB hosts could impact conclusions drawn from the integrated properties of galaxies observed at higher redshift (see Figure 8.8).

Several studies of this nature have been undertaken and reassuringly broadly support the conclusions drawn from either studies of the integrated light from the galaxies or the locations in broadband imaging. The locations of GRBs within their hosts are normally of low metallicity, even if the galaxy itself contains some higher-

metallicity regions. The stellar populations appear to be young, consistent with still containing their most massive stars. Future observations, possible with existing telescope upgrades to highly effective adaptive optics systems enabling Hubble-like resolution or with new extremely large telescopes, should make the spatially resolved approach possible across the full range of observed GRB redshifts, greatly strengthening the conclusions that are possible.

8.10 Comparative Properties of GRB Hosts with Other Core-collapse Events

Because of the very different selection techniques, a direct comparison of GRB host galaxies with other populations can be challenging. Outside of the very local universe, galaxy surveys suffer from significant incompleteness, especially for galaxies of low luminosity, so attempts to compare require extrapolation into regimes that are unknown. Indeed, even at low redshift, there may be significant evolution in the galaxy properties, since $z = 0.3$ converts to a look-back time of 3.5 Gyr, 25% of the age of the universe, and a time over which the star formation rate density has evolved significantly.

An alternative to making these comparisons is to attempt to compare GRB hosts with the host galaxies of transients selected via similar mechanisms, most logically core-collapse supernovae. In principle, core-collapse supernovae should arise from all stars with masses $> 8 M_{\odot}$, although this remains to be observationally confirmed, and very massive stars may collapse directly (Smartt 2015). Many such events are now located via blind surveys in which the sky is repeatedly tiled and transients are found. Hence, they are not preferentially located in massive galaxies, as used to be the case. By comparing these galaxies with those that host long GRBs, it should be possible to compare the environmental requirements for GRB production with the environmental properties for core-collapse supernova production directly, without the need for knowledge about, e.g., the galaxy luminosity function.

Such an attempt was first undertaken by Fruchter et al. (2006) using supernovae discovered as part of the GOODS survey that were well matched to the low-redshift GRB population and had *HST* optical imaging at the same wavelengths and spatial resolution. These observations showed that GRB hosts were systematically smaller and less luminous, had higher specific star formation rates (Svensson et al. 2010), and were more concentrated on their host galaxy light. Indeed, this was some of the first evidence for both a metallicity threshold and an origin in the most massive stars.

More recently, comparative studies have focused on the host galaxies of the hydrogen-poor superluminous supernovae (Lunnan et al. 2014, 2015; Angus et al. 2016; Perley et al. 2016b). Indeed, the hosts of the superluminous supernovae appear to share many properties with the GRB hosts, which is particularly relevant given the suggestion that similar central engines may be operating within each class of event (see Chapter 4). It seems likely that continued contrasting studies of astrophysical transient hosts can therefore provide strong clues to the progenitors, as well as uncovering potential links between classes of event.

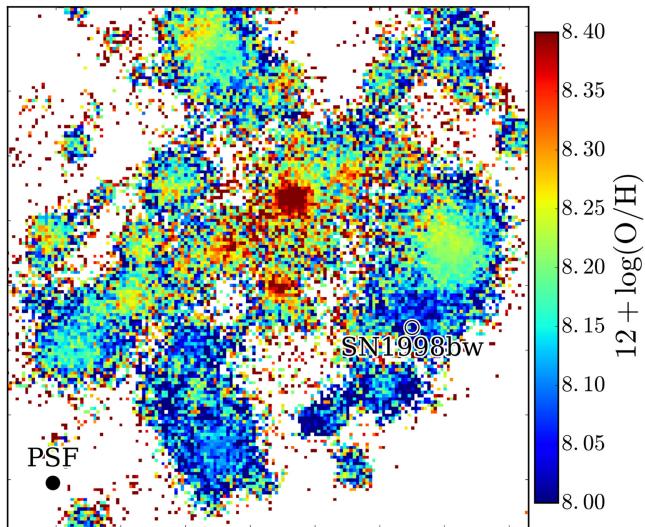


Figure 8.8. Host galaxy of the nearby GRB 980425, associated with SN 1998bw, taken from Krühler et al. (2017). Filled regions indicate where emission lines were visible for the determination of metallicity, while the scale for the metallicity is shown on the right-hand axis. It is clear that while the overall metallicity of the host galaxy is modest, with a central metallicity of $Z = 12 + \log O/H \sim 8.4$, the site of the GRB itself is significantly below this, consistent with a low-metallicity cutoff for GRB production. Reproduced with permission from Krühler et al. (2017) © ESO.

8.11 Summary

The host galaxies of long-duration GRBs have proved themselves to be exceptionally valuable tools for ascertaining the GRB progenitors. While the identification of supernovae ultimately tied them to the collapse of massive stars, the details of the stellar progenitors are perhaps better constrained by the environments in which they are found than they are from models of the supernova explosion itself. It is from these studies that we have firmly identified a metallicity bias in GRB production and hence know that such events are expected to be rare in galaxies like our own. These studies have also provided some of the first handles on the masses of the stars involved.

Beyond this, the use of GRB hosts to address questions of cosmological relevance has become increasingly clear in recent years. Regardless of their environmental biases, GRB hosts are a tool to probe the star formation history. Their afterglows provide one of the few routes of measuring galaxy metallicities at high redshift and, combined with host stellar masses inferred through imaging observations, can provide direct constraints on the buildup of stellar mass and metals across cosmic history. The dark GRBs probe obscured star formation directly. Indeed, this ability to identify galaxies based purely on the properties of the stars (not the presence of dust or starlight) is perhaps what makes GRBs most valuable in answering these questions.

References

- Angus, C. R., Levan, A. J., Perley, D. A., et al. 2016, *MNRAS*, **458**, 84
- Arabsalmani, M., Le Floc'h, E., Dannerbauer, H., et al. 2018, *MNRAS*, **476**, 2332
- Berger, E., Fox, D. B., Kulkarni, S. R., Frail, D. A., & Djorgovski, S. G. 2007, *ApJ*, **660**, 504
- Berger, E., Zauderer, B. A., Chary, R.-R., et al. 2014, *ApJ*, **796**, 96
- Blanchard, P. K., Berger, E., Fong, W., et al. 2017, *ApJL*, **848**, L22
- Bloom, J. S., Djorgovski, S. G., & Kulkarni, S. R. 2001, *ApJ*, **554**, 678
- Bloom, J. S., Kulkarni, S. R., & Djorgovski, S. G. 2002, *AJ*, **123**, 1111
- Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, *ApJ*, **670**, 928
- Campisi, M. A., Tapparello, C., Salvaterra, R., Mannucci, F., & Colpi, M. 2011, *MNRAS*, **417**, 1013
- Caraveo, P. A., Mignani, R. P., Tavani, M., & Bignami, G. F. 1997, *A&A*, **326**, L13
- Cenko, S. B., Kelemen, J., Harrison, F. A., et al. 2009, *ApJ*, **693**, 1484
- Christensen, L., Hjorth, J., & Gorosabel, J. 2004, *A&A*, **425**, 913
- Conselice, C. J. 2003, *ApJS*, **147**, 1
- Conselice, C. J., Vreeswijk, P. M., Fruchter, A. S., et al. 2005, *ApJ*, **633**, 29
- Fruchter, A. S., Levan, A. J., Strolger, L., et al. 2006, *Natur*, **441**, 463
- Fynbo, J. P. U., Jakobsson, P., Møller, P., et al. 2003, *A&A*, **406**, L63
- Gorosabel, J., Pérez-Ramírez, D., Sollerman, J., et al. 2005, *A&A*, **444**, 711
- Graham, J. F., & Fruchter, A. S. 2013, *ApJ*, **774**, 119
- Graham, J. F., & Fruchter, A. S. 2017, *ApJ*, **834**, 170
- Groot, P. J., Galama, T. J., van Paradijs, J., et al. 1998, *ApJL*, **493**, L27
- Hatsukade, B., Ohta, K., Endo, A., et al. 2014, *Natur*, **510**, 247
- Hjorth, J., Malesani, D., Jakobsson, P., et al. 2012, *ApJ*, **756**, 187
- Jakobsson, P., Hjorth, J., Malesani, D., et al. 2012, *ApJ*, **752**, 62
- James, P. A., & Anderson, J. P. 2006, *A&A*, **453**, 57
- Krühler, T., Kuncarayakti, H., Schady, P., et al. 2017, *A&A*, **602**, A85
- Krühler, T., Malesani, D., Fynbo, J. P. U., et al. 2015, *A&A*, **581**, A125
- Krühler, T., Malesani, D., Milvang-Jensen, B., et al. 2012, *ApJ*, **758**, 46
- Larsson, J., Levan, A. J., Davies, M. B., & Fruchter, A. S. 2007, *MNRAS*, **376**, 1285
- Levan, A., Fruchter, A., Rhoads, J., et al. 2006, *ApJ*, **647**, 471
- Lunnan, R., Chornock, R., Berger, E., et al. 2014, *ApJ*, **787**, 138
- Lunnan, R., Chornock, R., Berger, E., et al. 2015, *ApJ*, **804**, 90
- Lyman, J. D., Levan, A. J., Tanvir, N. R., et al. 2017, *MNRAS*, **467**, 1795
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, **408**, 2115
- McGuire, J. T. W., Tanvir, N. R., Levan, A. J., et al. 2016, *ApJ*, **825**, 135
- Metzger, M. R., Djorgovski, S. G., Kulkarni, S. R., et al. 1997, *Natur*, **387**, 878
- Michałowski, M. J., Castro Cerón, J. M., Wardlow, J. L., et al. 2016, *A&A*, **595**, A72
- Michałowski, M. J., Kamble, A., Hjorth, J., et al. 2012, *ApJ*, **755**, 85
- Perley, D. A., Hjorth, J., Tanvir, N. R., & Perley, R. A. 2017, *MNRAS*, **465**, 970
- Perley, D. A., Levan, A. J., Tanvir, N. R., et al. 2013, *ApJ*, **778**, 128
- Perley, D. A., Perley, R. A., Hjorth, J., et al. 2015, *ApJ*, **801**, 102
- Perley, D. A., Krühler, T., Schulze, S., et al. 2016a, *ApJ*, **817**, 7
- Perley, D. A., Quimby, R. M., Yan, L., et al. 2016b, *ApJ*, **830**, 13
- Perley, D. A., Tanvir, N. R., Hjorth, J., et al. 2016c, *ApJ*, **817**, 8
- Raskin, C., Scannapieco, E., Rhoads, J., & Della Valle, M. 2008, *ApJ*, **689**, 358

- Rol, E., Wijers, R. A. M. J., Kouveliotou, C., Kaper, L., & Kaneko, Y. 2005, *ApJ*, 624, 868
Sahu, K. C., Livio, M., Petro, L., et al. 1997, *Natur*, 387, 476
Savaglio, S., Glazebrook, K., & Le Borgne, D. 2009, *ApJ*, 691, 182
Schady, P., Mason, K. O., Page, M. J., et al. 2007, *MNRAS*, 377, 273
Schulze, S., Chapman, R., Hjorth, J., et al. 2015, *ApJ*, 808, 73
Smartt, S. J. 2015, *PASA*, 32, e016
Stanway, E. R., Levan, A. J., Tanvir, N. R., Wiersema, K., & van der Laan, T. P. R. 2015, *ApJL*, 798, L7
Starling, R. L. C., Vreeswijk, P. M., Ellison, S. L., et al. 2005, *A&A*, 442, L21
Svensson, K. M., Levan, A. J., Tanvir, N. R., Fruchter, A. S., & Strolger, L.-G. 2010, *MNRAS*, 405, 57
Tanvir, N. R., Mackey, A. D., Ferguson, A. M. N., et al. 2010, *MNRAS*, 422, 162
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, 613, 898
Vergani, S. D., Salvaterra, R., Japelj, J., et al. 2015, *A&A*, 581, A102
Wainwright, C., Berger, E., & Penprase, B. E. 2007, *ApJ*, 657, 367

Gamma-Ray Bursts

Andrew Levan

Chapter 9

Multimessenger Astronomy

9.1 From Multiwavelength to Multimessenger Astronomy

For most of human history, astronomy has been conducted at visible wavelengths, predominantly with the naked eye. In the 20th century, remarkable progress was made with the opening of the electromagnetic spectrum to astronomical observations. The advent of multiwavelength astronomy was transformational, and many insightful and important discoveries were made, from radio observations of the cosmic microwave background (CMB), to the discovery of pulsars, to the identification of stellar and supermassive black holes with X-ray observations, to the discovery of GRBs. Indeed, since the 1975 Nobel Prize was awarded to Martin Ryle and Antony Hewish, a further five Nobel prizes have been awarded for observations outside of the optical window.¹ The frontier is now to move beyond multiwavelength astronomy and into the era of multimessenger astronomy, where nonphotonic messengers are used to carry information inaccessible to any wavelength of electromagnetic light.

This “new” information is valuable because there are important regions of the universe that are far too optically thick for the direct detection of light at any wavelength. This may include the early universe before recombination, the cores of collapsing stars, and the remnants of compact object mergers. Beyond this, even when electromagnetic light is emitted, the combination of the light with an additional messenger provides a new route to distinguish the physical mechanisms at play. For example, the detection of neutrinos from SN 1987A confirmed that electron capture was occurring and that the bulk of the energy of a supernova is carried away by neutrinos (Hirata et al. 1987). The detection of a short GRB with a gravitational-wave detection of a merging neutron star binary immediately answers the question of the progenitor of that short burst (Abbott et al. 2017e).

¹ These are 1978 for the discovery of the CMB, 1993 for the binary pulsar, 2002 for the discovery of cosmic neutrons and the development of X-ray astronomy, 2006 for the spectra and anisotropy of the CMB, and 2017 for the detection of gravitational waves.

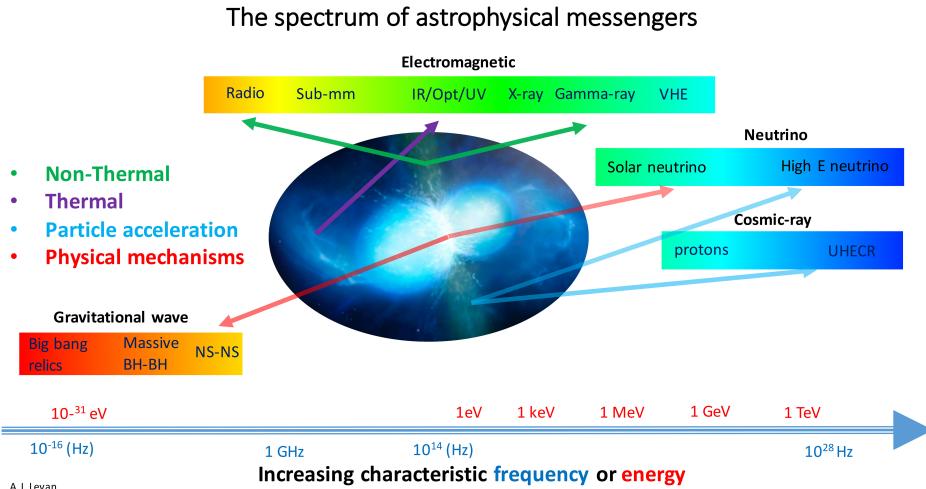


Figure 9.1. Energy spectrum of currently accessible astrophysical messengers. The vast majority of astronomical observations are carried out in electromagnetic light, and the multiwavelength revolution transformed 20th-century astronomy. We are now entering a new period of discovery in the so-called multimessenger era, where these electromagnetic signals are combined with additional, nonphotonic messengers, including gravitational waves, neutrinos, and high-energy cosmic rays (although the latter are massive and charged and so cannot be directly traced back to an individual source). These additional diagnostics provide information about the behavior of mass and particle acceleration. Combined, they enable highly complete physical pictures to be built around many of the most extreme objects thought to exist in nature. The energy/frequency range spanned by these observations is extreme. The lowest-frequency gravitational waves can probe structures in the early universe that have oscillated through only a handful of periods in the Hubble time. In contrast, the most energetic cosmic rays have energies of 100 EeV, an essentially macroscopic energy. (Inset figure credit: University of Warwick/Mark Garlick).

The new dimension enabled by multimessenger observations is shown in Figure 9.1, which shows the multimessenger spectrum. This carries information from very low frequency gravitational waves, which oscillate only a few times in the age of the universe, to the very high energy neutrinos and cosmic rays, with energies of TeV or frequencies of 10^{28} Hz. For an extreme event, such as a binary neutron star merger, it is plausible that electromagnetic light, gravitational-wave, and neutrino signals are all generated with sufficient strength for detection with current (or at least near-future) instrumentation. The combination of γ , gravitational-wave, and ν signals would provide information on the motions of mass, thermal and nonthermal processes, particle acceleration, and more, perhaps enabling the full details of the result of the merger to be described for the first time.

This chapter outlines the promise of multimessenger astronomy with both gravitational waves and neutrinos, with a particular focus on the relationships to GRBs as multimessenger sources.

9.2 Gravitational Waves

Gravitational waves were first predicted as a consequence of general relativity in 1915. The principles behind their formation lie in the distortion of spacetime by

masses. As these masses accelerate, they must move spacetime with them, but such propagation cannot be instantaneous, so gravitational waves are formed. More specifically, such waves do not arise from all accelerating systems but only from those that accelerate in a nonspherically symmetric manner, so orbiting systems are prime sources of gravitational waves. Once created, the waves are not impeded by the presence of any matter surrounding the source by travel at the speed of light across the universe. Since gravitational waves are directly ripples in spacetime, they manifest themselves as tiny differences in the distance between any two points in space, and these distances oscillate as the waves pass through. Since spacetime is exceptionally stiff (i.e., the degree of curvature, even for very large masses, is very limited), the amplitude of these waves is small and thus only becomes important for very dense, compact sources.² Like electromagnetic waves, gravitational waves carry two polarizations. Unlike the electric and magnetic components in electromagnetic waves that are perpendicular to each other, the gravitational-wave polarizations lie at 45° and are often referred to as the + and \times polarization. Since the waves themselves carry energy, this is sourced from the accelerating sources. In the case of orbiting systems, the gravitational waves come at the expense of orbital energy, so the orbit of a system gradually shrinks. For systems with sufficient mass that are initially close, this forces a merger of the two components in less than the Hubble time. Direct evidence for this can be seen in galactic systems containing two neutron stars, in which the change of period due to gravitational-wave radiation can be directly measured and is in remarkable agreement with the expectations of general relativity (Hulse & Taylor 1975; Weisberg & Taylor 2005); see Figure 9.2.

The detailed mathematics and physics behind gravitational waves and their origin and generation is beyond the scope of this text but is covered in varying degrees of detail in many general-relativity works. Here we instead look at the predictions for gravitational waves that are likely to arise from the progenitors of gamma-ray bursts and discuss the prospects for joint detection.

9.2.1 Gravitational-wave Detectors and Localization

Gravitational waves are detected by the measurement of minute changes in length between objects due to the passage of the wave. This is typically classified as a strain amplitude, or $\Delta L/L$. This dimensionless quantity implies that more extended lengths of the detector enable more sensitive searches, assuming the change in length can be measured to the same precision. Early attempts at detection relied on an approach known as a resonant bar (often called a Weber bar, after Joseph Weber, who pioneered their use), in which test masses, typically made of aluminum, were placed in different locations. These bars had set resonant frequencies, and if they were vibrated at these frequencies, they would provide a signal of a gravitational-wave source whose emission was sweeping through this band. This frequency was in the range of 1500 Hz and would correspond to the final fraction of a second of a

²Indeed, the density is much more important than the proximity. All of our day-to-day movements create gravitational waves, but they cannot be detected, even at very small distances.

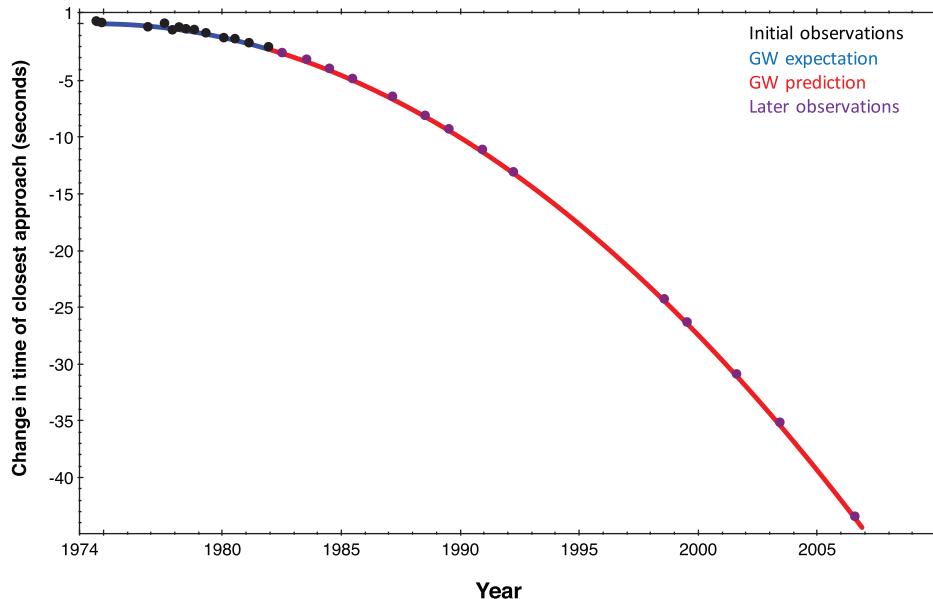


Figure 9.2. Change in periastron of the Hulse–Taylor pulsar. The observations initially reported and compared to the expectations of gravitational-wave radiation are shown in black (with the prediction in blue). Beyond this point, it was possible to make strong predictions of the emission (red), which subsequently proved extremely accurate (magenta). Data from Weisberg & Taylor (2005).

compact object merger. However, the strain sensitivities reached by these detectors were not sufficient to achieve the expected strains for the majority of astrophysical systems. Indeed, it was perhaps somewhat surprising that these early observations produced claims of detection (Weber 1969), which subsequently appeared unlikely or were even directly ruled out by later, more sensitive observations both with resonant-bar detectors in other facilities and by new laser interferometers.

The gold standard for gravitational-wave observations today is the use of laser interferometers, such as the Laser Interferometer Gravitational wave Observatory (LIGO) detector, operated in Hanford in Washington state and Livingston in Louisiana; the VIRGO interferometer in Italy; the Kamioka Gravitational Wave Detector (KAGRA) system in Japan; and the GEO600 interferometer in Germany. These systems are essentially scaled-up Michelson–Morely interferometers with exceptionally powerful lasers and extremely stable setups (see Figure 9.3). They have path lengths of up to 4 km (in the case of LIGO) but recycle their laser power, such that the light bounces back up and down the arms of the interferometer many times, creating a much longer effective interferometer length.³ This light is then combined and destructively interfered so that no light is seen at the detector. In principle, flashes of light that are then seen from the system are indicative of changes in the path length of one of the arms relative to the other, which brings the signal out of

³This also, of course, increases the travel time of the light, but the travel time is short compared to the frequency of the sources that the interferometers search for, so this does not cause a problem.

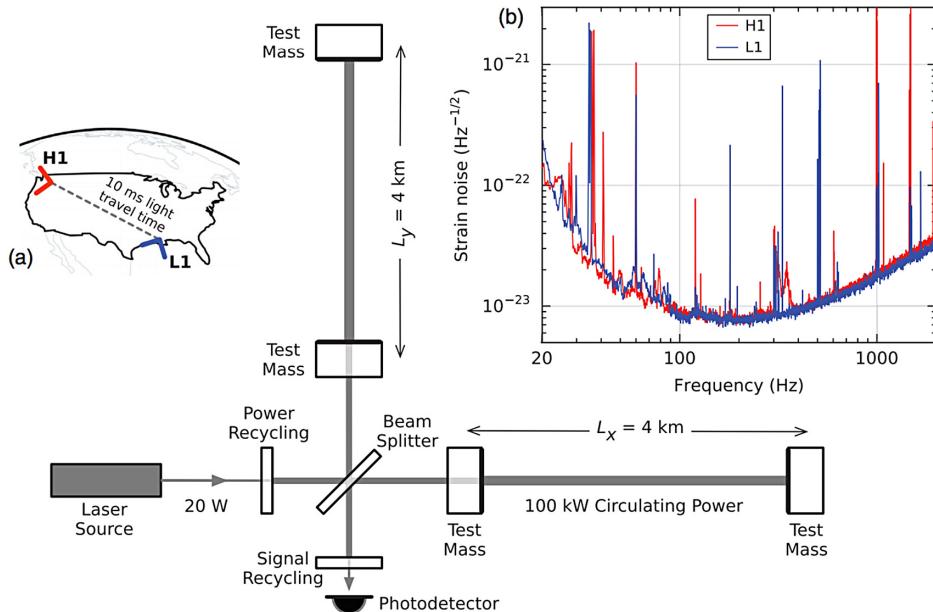


Figure 9.3. Schematic layout of the LIGO detectors in Hanford (Washington state) and Livingston (Louisiana) from Abbott et al. (2016d). Panel (a) shows the relative geographical locations; this provides a time-of-flight measurement for gravitational waves, and is the route that enables a degree of localization. Panel (b) shows the relative strain sensitivity as a function of the frequency of a gravitational wave. An inspiral signal will sweep through this band, increasing in amplitude up until the point of merger (more massive systems merge at lower frequency). The shape of the sensitivity curve is dominated by seismic noise at low frequency and shot noise in the lasers at high frequency. The main panel shows the layout of the interferometer, in which two 4 km arms recycle a 20 W laser approximately 5000 times to create 100 kW of circulating power. The arms of the interferometer recycle the laser off two seismically isolated test masses and then destructively interfere with the output, resulting in no output signal in the absence of gravitational waves. Reprinted figure with permission from Abbott et al. Copyright (2016) by the American Physical Society. CC BY 3.0

perfect cancellation. The intensity of the light seen can then be converted into a change in length and mapped accurately over time, providing a time-variable measurement of the oscillation of the interferometer as the signal passes through it. These length changes can either be quantified in terms of the relative strength of the wave (for a so-called gravitational-wave burst, which shows no particular structure) or compared to the pattern expected from general relativity, for example, a merging double compact object binary. In practice, both template match and burst searches are conducted on the data stream using a variety of different techniques.⁴

Gravitational waves cannot be focused, so localization from a single detector is not possible. Instead, triangulation must be used to provide positions for any

⁴The details of gravitational-wave instrumentation and data analysis are complex and beyond the level of this text; for further information, the reader is directed to LIGO Scientific Collaboration et al. (2015) and Abbott et al. (2016e).

detected sources. To first order, this can simply be done via the time delays between detection at different observatories, with delays of a few milliseconds introduced by gravitational-wave travel times (at c). In this case, two detectors can provide an annulus across the sky and a third a more constrained sky region. Since the detectors are on the Earth, which is a curved surface, different detectors also have different orientation to any source in the sky. Since the gravitational-wave polarization and strain are a function of the angle to the detector, this can break degeneracies in the circle provided by only two detectors, constraining a smaller sky region from which the source may arise.⁵ For two detectors, the typical sky regions are hundreds of square degrees (90% containment), but for three detectors, they can be reduced to tens of square degrees (Fairhurst et al. 2011).

Searches for electromagnetic counterparts may be based on time coincidence (e.g., in the case of a gamma-ray burst), but for detailed studies of the source, they will rely on pinpointing the source on the sky to arcsecond levels. Hence, the viability of counterpart searches is related to the area of sky to which a given source can be localized.

9.2.2 Gravitational-wave Searches and Detection

The strains expected for most sources are extremely small (see below) and typically lie in the region of $\Delta L/L < 10^{-22}$. Because of this, several generations of interferometer failed to find any convincing astrophysical signals to ever more impressive limits. The initial generation of LIGO observatories reached peak strain noises of a few $\times 10^{-23}$ at frequencies of 100–500 Hz, suggesting that they could have detected sources with strains smaller than 10^{-21} at high significance, but they did not detect any convincing gravitational-wave sources.

This has finally changed with the advent of the most recent generation of gravitational-wave interferometers, which have been upgraded from the previous generation and can now reach strain noises better than 10^{-23} (see Figure 9.3). They attain approximately this sensitivity across a range of frequencies from 100 to 1000 Hz. At the low-frequency end, the sensitivity rapidly degrades because of seismic interference, while at high frequencies, the sampling rate of the laser (so-called shot noise) is the limiting factor simply because of the errors on counting statistics on the individual photons.

These detectors have finally reached the sensitivity where astrophysical detections were expected. Indeed, the first detection occurred just before the first major science run with Advanced LIGO commenced. On 2015 September 14, a strong signal was detected (Abbott et al. 2016d). It was visible in the band for a fraction of a second and then stopped, consistent with the merger of two relatively massive objects. Subsequent analysis suggested that the signal arose from the merger of two black holes with initial masses of 36^{+5}_{-4} and $29^{+4}_{-4} M_{\odot}$ to form a new black hole with a total mass of $62^{+4}_{-4} M_{\odot}$. The remaining $\sim 3 M_{\odot}$ of mass was lost to gravitational waves.

⁵ Within the curvature of the Earth, the two LIGO detectors are aligned in the same direction, so they provide limited polarization information.

The inspiral waveforms for the merging black holes, as seen in the two LIGO detectors in Hanford and Livingston, are shown in Figure 9.4. The increasing frequency and amplitude of the merger is clearly seen and results in the so-called gravitational-wave chirp.

These merging binaries were expected to be the main source of detectable gravitational waves (see below for more details), providing much stronger signals than possible via other routes. However, the detection of black holes much more massive than those seen within the Milky Way, for example, within X-ray binaries, was at first sight more surprising. However, the strain scales with the total mass involved, so such massive systems can be seen to larger distances and are not necessarily common. Further observations of black hole–black hole binaries have been forthcoming with LIGO (Abbott et al. 2016a, 2016b, 2017c) and, more

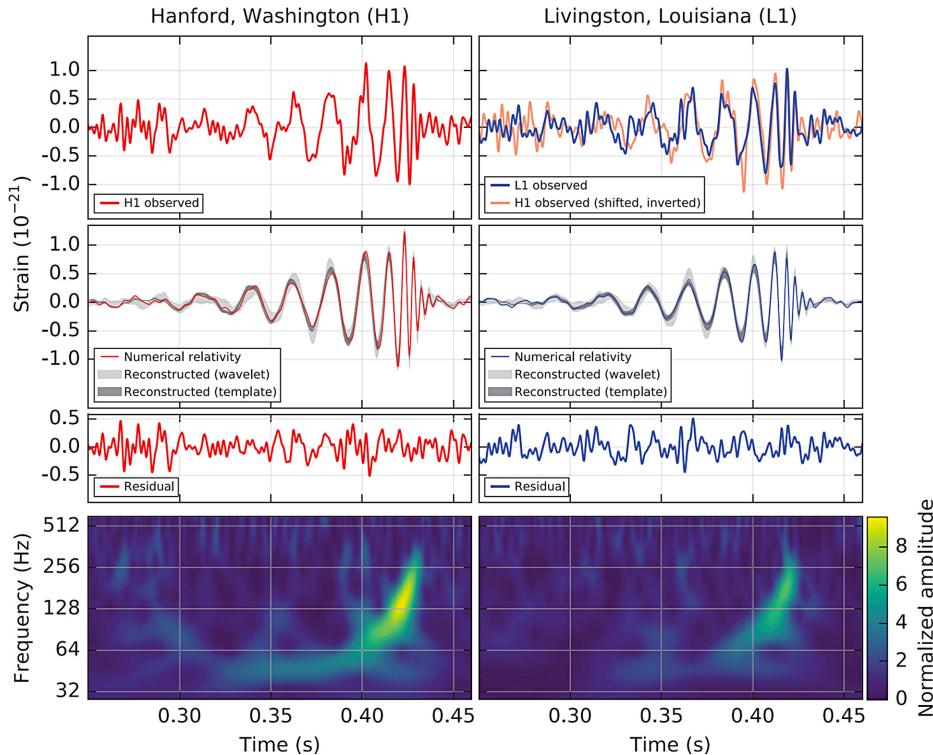


Figure 9.4. Detection of the first gravitational-wave source (GW 150914) in the two LIGO detectors (Abbott et al. 2016d). The top left panel shows the signal in the Hanford detector, while the top right panel shows the Livingston detector, as well as the Hanford signal overlaid, appropriately shifted for the different detector sites. The middle panels show the merger templates from numerical relativity, along with the residuals to the fit. The bottom panels show the characteristic chirp signal of increasing frequency and amplitude. In this case, the large black hole masses, 36^{+5}_{-4} and $29^{+4}_{-4} M_{\odot}$, create a high-amplitude signal, although there are relatively few orbits in the LIGO band, as the merger happens at a relatively low frequency. Reprinted figure with permission from Abbott et al. Copyright (2016) by the American Physical Society. CC BY 3.0

recently, with VIRGO as well (Abbott et al. 2017d). These confirm that relatively massive black holes are common, but none have been as massive as the first event.

Black holes are likely the outcome of stellar evolution only for massive stars, thus black hole–black hole systems created through stellar evolution alone should be rare. It is possible (or even likely) that some of those observed to date have in fact been created dynamically in dense clusters (Rodriguez et al. 2016). For lower-mass stars, neutron star remnants are likely, and systems containing neutron stars (either neutron star–black hole or neutron star–neutron star) should have volumetric rates much higher than black hole–black hole systems. They are more difficult to detect owing to the lower mass and resulting smaller strain. After five detections of black hole–black hole mergers, a neutron star binary was finally discovered in 2017 August (Abbott et al. 2017e). Observations of this system and searches for and implications of electromagnetic counterparts are described in more detail below.

9.3 Sources of Gravitational-wave Emission

There are several possible sources of gravitational-wave emission, but they can be reduced to four general categories, as follows.

- **Persistent sources**—Persistent sources of gravitational-wave emission are those for which the frequency is constant (or changes slowly). At higher frequencies (e.g., those to which LIGO is sensitive), they are likely to be related to the spins of compact stars—for example, pulsars or magnetars—in which any variation from spherical symmetry within the neutron star will become apparent as a gravitational-wave signal as the star rotates (Abbott et al. 2017f). At lower frequencies, compact binary stars such as neutron star–neutron star or even white dwarf–white dwarf binaries are likely to be dominant foreground sources of gravitational waves (Kupfer et al. 2018).
- **Inspiral sources**—The strongest expected signals of gravitational waves arise from compact binary stars (neutron star–neutron star, neutron star–black hole, black hole–black hole) in their final phases (Abadie et al. 2010). During this phase, the orbital separation is rapidly decreasing and the frequency of the binary is rapidly increasing, along with the gravitational-wave strength. The precise pattern in frequency and the point of merger depends sensitively on the masses of the two components, such that inspirals can be identified via template matching of a range of possibilities against the observed data.
- **Bursting sources**—These refer to sources where gravitational waves are produced that are incoherent. They will result in net power in the detector but will not match any expectation for either a persistent source (e.g., of fixed frequency) or an inspiral. Gravitational-wave bursts are expected from core-collapse events, but their strength depends strongly on the asymmetry of the collapse (Kotake et al. 2006).
- **Stochastic sources**—For the sake of completeness, it is also worth noting that there is the possibility of stochastic gravitational-wave emission. This would originate from the early stage of the universe and perhaps provide information about cosmological inflation. In particular, it can provide information

from an era prior to the formation of the CMB. This is point in time that electromagnetic observations cannot probe because the universe was opaque. Stochastic gravitational waves are not related to individual sources and therefore not a source for multimessenger astronomy.

Below, the expected properties of these different kinds of gravitational-wave sources and their prospects as multimessenger detections are outlined in more detail.

9.3.1 Persistent Sources: Compact Object Binaries and Neutron Stars

Persistent gravitational-wave sources have the advantage that their counterparts are already well known. There is good evidence for gravitational-wave radiation in the shrinking orbits of double neutron stars observed in the Milky Way today, and it is expected that a population of white dwarf–white dwarf binaries dominates the low-frequency regime (10^{-6} Hz; Nelemans 2009) that is not accessible to ground-based detectors because of seismic noise but is the target of future space-borne projects such as the *Laser Interferometer Space Antenna* (*LISA*).⁶ The presence of these sources is well known, and their gravitational-wave emission can be measured from the changes in their period without the need for direct detection.

However, there are expected sources of higher-frequency emission that may ultimately be detected by the current generation of ground-based interferometers. The prime source is likely to be measurements of gravitational waves from rotating neutron stars (Aasi et al. 2014). Any deviation from spherical symmetry will result in the emission of gravitational waves as the neutron star spins. Because of this, “mountains” on neutron stars can be measured directly and provide constraints on the equation of state of the neutron star itself.

The challenge with such work is that most neutron stars have relatively long spin periods, so the frequency of emission is not well matched to the peak sensitivity of the LIGO/VIRGO instruments, which begins primarily at around 20 Hz (spin periods of 0.05 s). Hence, the best sources are those that are spun up to become millisecond pulsars. The analysis necessary to interpret persistent sources differs from that used for transient sources, since it is unlikely that individual rotations will yield the required strain amplitudes for detection. Instead, attempts at stacking or cross-correlation are necessary.

To date, there have been no detections of individual rotating neutron stars. However, the limits possible with the current generations of detectors are now reaching strain amplitudes of 10^{-25} and are within a factor of a few of those predicted under certain models (Abbott et al. 2017f), so it is possible that the current generation of detectors, with a sufficiently long baseline, could provide such detections.

9.3.2 Inspiral Sources: Compact Object Binaries

Compact object binary inspirals are the prime transient sources expected to be observed by gravitational-wave observatories. Their inspiral from separations of

⁶<http://sci.esa.int/lisa/>

$\sim R_\odot$ can take place in less than a Hubble time and is driven by the emission of gravitational-wave radiation. The amplitude of this radiation and its frequency increases as the binary shrinks, and the final minutes of the inspiral of the neutron star binary are visible in the LIGO frequency range. The measurement of the frequency evolution of the orbit while in the band pinpoints the so-called chirp mass of the system,

$$\mathcal{M} = (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}, \quad (9.1)$$

and provides some (weaker) constraints on the masses of the individual components. Lower-mass systems (e.g., those containing neutron stars) remain in band for longer and merge at a higher frequency than higher-mass systems (those containing black holes).

Routes to the formation of compact object binaries are described in Chapter 6, but briefly, they can be formed via binary evolution of massive stars in which the binary survives the supernova creating each compact object and where the separation is shrunk to the scale of solar radii via common envelope evolution. Alternatively, it is also possible to form compact object binaries dynamically in dense environments such as star clusters close to galactic nuclei or in globular clusters. Indeed, it has been suggested that the black hole–black hole binaries are preferentially formed in such environments (Rodriguez et al. 2016), where sequential mergers can build up the higher masses. However, such a route is not unambiguous, and it is also possible to form massive black holes via binary channels, especially at lower metallicity (de Mink & Mandel 2016; Eldridge & Stanway 2016).

The appeal of compact object binaries as gravitational-wave sources arises not only because they are the strongest likely sources (and indeed, the only sources yet detected) but also because there are strong expectations of electromagnetic signals that should be visible from them, in particular, if the merger involves a neutron star.

9.3.3 Bursting Sources: Core-collapse Events

Long-duration GRBs arise from core-collapse supernovae. In principle, a core collapse is a spherically symmetric event, in which case a long GRB would not create any visible gravitational-wave emission. However, it is unlikely that GRB supernovae (or indeed any supernovae) are actually symmetric. Asymmetries within the star, created either by rotation or by off-center locations for the formation of the supernovae can produce a wide range of different asymmetries, and these may well be responsible for the neutron star kicks that we observe. Indeed, the presence of apparent natal kicks to neutron stars is in itself strong evidence for the existence of asymmetry in the collapse process, since an entirely spherical collapse should yield a neutron star at rest (since the mass infall during core collapse would be the same in all directions). Since supernovae are common, gravitational-wave emission from supernovae would appear promising. However, the kicks imparted to neutron stars actually imply relatively little asymmetry in the collapse process. Asymmetries of only around 1% can easily result in neutron star velocities of hundreds of km s⁻¹, and this may mean that core-collapse supernovae do not create power

gravitational-wave emission. Indeed, one can estimate the expected strength of the gravitational-wave emission from these events as a function of the degree of asymmetry within them. Unlike compact binaries (see below), where the gravitational-wave signal is straightforward to compute, in core-collapse events, it depends on the details of the collapse process. Perhaps the most optimistic scenario is that the core of the collapsing star fragments as it collapses, providing clumps of material with nuclear density (effectively mini neutron stars) that coalesce under the combined effects of gravitational radiation and dynamical friction (Davies et al. 2002). In this scenario, the masses of the objects could be a considerable fraction of the neutron star mass, so the horizon for the events could be comparably large. However, simulations of core-collapse events do not show this degree of fragmentation, so it does not seem likely that this mechanism operates in the majority of systems.

Although it is a model-dependent statement, it seems likely that core-collapse events will only be visible to the current generation of gravitational-wave detectors should they arise within the Milky Way or, plausibly, the Magellanic Clouds.

9.4 Gravitational-wave Horizons

A crucial characterization of a gravitational-wave experiment, especially in the era where a paucity of direct detections exists, is the effective horizon (or volume) that it probes. This is a function of the signal being searched for. Generally, horizons are considered for merging compact objects, since these yield waveforms where the strain sensitivity can be converted into a horizon scale.

For electromagnetic waves, the flux or flux density is normally measured. This decreases as $1/d^2$, where d is the distance, this relationship is the standard inverse square law for light. For gravitational waves, the strain decays as $1/d$. Hence, a detector that is a factor of 10 more sensitive than its predecessor can see a factor of 10 further, rather than a factor of ~ 3 for electromagnetic waves. Since the number of possible sources scales with the volume enclosed within the horizon, this corresponds to a factor of ~ 1000 in the rate of detections. It is this feature that means that it is possible to evolve rapidly from a regime in which detections are not expected into one where they are routine.

The strength of the two gravitational-wave polarizations of a binary is given by Korol et al. (2017) as

$$h_+(t) = \frac{2(GM)^{5/3}(\pi f_s)^{2/3}}{c^4 d} (1 + \cos^2 i) \cos 2\Phi(t) \quad (9.2)$$

and

$$h_\times(t) = -\frac{4(GM)^{5/3}(\pi f_s)^{2/3}}{c^4 d} \cos i \sin 2\Phi(t), \quad (9.3)$$

where $\mathcal{M} = (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}$ is the chirp mass of the system, $\Phi(t) = \Phi_0 + \pi f_s t$ is the orbital phase, and i is the inclination of the binary orbital plane with respect to the

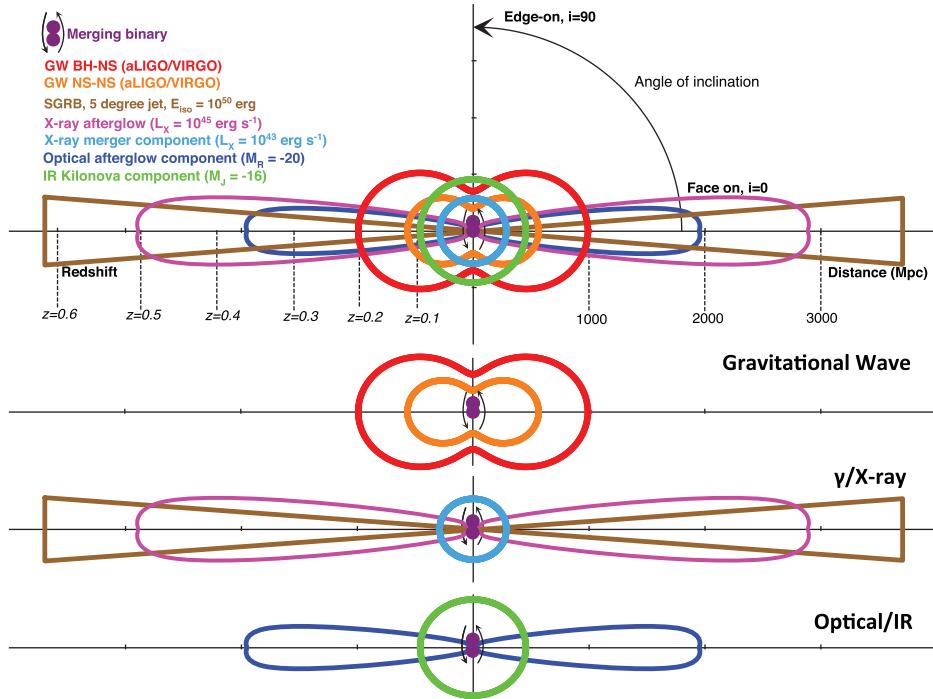


Figure 9.5. Horizons for the detection of gravitational waves and associated electromagnetic counterparts from Levan et al. (2016). With permission of Springer. CC BY 4.0

line of sight. The $\Phi(t)$ term is time variable and its sine/cosine can be set to unity to represent the maximum strain amplitude for a given set of (evolving) binary parameters.

It can also be seen that the strength of the signal is dependent on the orientation of the binary with respect to the observer. Gravitational waves that are observed face-on ($i = 0$) provide much stronger signals. In particular, $h_+(t)$ is a factor of two larger than for $i = 90$, while $h_\times(t)$ vanishes for $i = 90$. The total strain can be given as $h^2 = h_+^2 + h_\times^2$ and, for a given limit, can be converted into an effective horizon, shown in Figure 9.5. The gravitational-wave horizon is, therefore, a factor of ~ 2 larger for face-on systems. The true horizon depends not only on the source orientation to Earth but also to the detectors, which have an antenna pattern with varying sensitivity depending on the sky position. Therefore, rather than express horizons as directional numbers, it is most common to give sky-averaged distances. For the Advanced LIGO detectors, the design sensitivity is a sky-averaged horizon of a few hundred Mpc for a binary neutron star ($2 \times 1.4 M_\odot$) and can be scaled for more massive systems. Indeed, black hole–black hole mergers have already been seen out to $z > 0.2$ (> 1000 Mpc).

9.5 Prospect for Joint Detections

Gravitational waves alone provide a wealth of astrophysical information, but this is greatly enhanced by the addition of electromagnetic (or indeed neutrino)

information. In particular, while gravitational waves provide good measurements of compact object masses, as well as potentially the radii of neutron stars (since the merger signal looks different if it ends at the radius of a neutron star or the Schwarzschild radius of a black hole, so potentially the deformability measurements of the neutron star, which become visible as asymmetry in the waveform, enable a constraint on the neutron star equation of state and hence radius), they do not by themselves provide a link back to directly observed systems and signals that are emitted across the electromagnetic spectrum. As examples, a gravitational-wave burst source may arise in a core-collapse event, but since the burst does not yield easily to modeling, we do not know if this is correct; we are unable to determine if the collapse led to an outward shock or direct collapse to a black hole or what kind of supernova may have been created. For merging binaries, the gravitational-wave observations provide precise masses and luminosity distances—so-called standard sirens (Schutz 1986)—but these cannot be tied to redshifts to create a Hubble diagram (to measure the Hubble constant) without electromagnetic signals. The gravitational-wave observations will also provide a relatively bias-free view of the rates of compact object mergers, but they will not provide information on the heavy-element yield within these mergers and thus cannot alone determine if they are the dominant site for the synthesis of the heaviest elements. In each of these cases, and indeed many others, there is great scientific return from the detection of both gravitational-wave and electromagnetic signatures. The possible routes to the identification of these signatures, their likelihood of detection, and the scientific returns are outlined below.

9.5.1 Gamma-Ray Bursts

The GRBs are promising routes for the detection of electromagnetic counterparts because they are bright events, readily identified in all-sky monitors (such as the IPN) as well as deeper surveys covering smaller fractions of the sky, such as *Swift* (approximately one-sixth of the sky) or *Fermi* (only limited by Earth blocking). Since GRBs are also relatively rare events, with detection rates of a few per week, their frequency is sufficiently low that the coincident detection of a GRB and gravitational-wave trigger within a few seconds of each other has a low chance of probability. This means that time information alone can enable the association of a GRB with a gravitational-wave transient, even in the absence of any positional information, as was the case for GW 170817 and GRB 170817A (see below). Given the challenges of searching gravitational-wave error boxes, this is extremely appealing. Indeed, the error boxes derived from these gamma-ray detectors can frequently be much smaller than those from the gravitational-wave event alone, providing a far easier search for afterglows or host galaxies. Finally, the precise timing of GRB triggers also narrows the windows for gravitational-wave searches, lowering the associated background. This means that even in the absence of a gravitational-wave trigger, the data stream can be analyzed only around the time of the GRBs, increasing the effective sensitivity (and hence distance range).

9.5.2 Kilonovae

Kilonovae are radioactively driven transients powered by the synthesis of heavy elements in the neutron-rich ejecta from a binary merger (Li & Paczyński 1998; Metzger & Berger 2012; Barnes & Kasen 2013). They are described in more detail in Chapter 6 but are outlined here for completeness. The principle of their production is that within a merger, not all of the material is retained by a central compact object. A small quantity, $\sim 0.01 M_{\odot} < M_{\text{eject}} < 0.1 M_{\odot}$, of material is ejected from the system, either via a polar wind or through expulsion in tidal tails. As this material decompresses from the neutron star, it begins nucleosynthesis, and the low proton (or electron) fraction, often described as Y_e , enables multiple rapid neutron captures (with a timescale for neutron capture that is less than the decay time for a given isotope) that build up r -process elements. The subsequent (mainly) β -decays of these isotopes enable the creation of ever heavier elements, as well as providing a heat source within the ejecta.

Since many of the heaviest elements produced, in particular the lanthanides, have high opacities to optical light, it is expected that kilonovae should be very red (Barnes & Kasen 2013; Berger et al. 2013; Tanvir et al. 2013), at least on timescales beyond a few days when lighter elements may dominate. Therefore, they should be straightforward to disentangle from other transient sources. In principle, identifying a kilonova within a gravitational-wave error box is possible because it will be (i) new, compared to preexisting imaging; (ii) faint, compared to a supernova at peak; and (iii) red.

9.5.3 Orphan Afterglows

While GRBs are appealing routes to the identification of compact object mergers, they are likely only visible to observers who view down the jet axis, and their visibility to off-axis observers depends sensitively on the structure and width of the associated relativistic jet. However, as jets slow down by plowing into the interstellar medium, they also expand sideways and eventually become closer to spherical explosions. This means that observers initially outside of the jet can begin to observe the resulting afterglow. These are often referred to as orphan afterglows, in recognition of the fact that they arise without a prompt gamma-ray trigger. To date, there have been a handful of candidate orphan afterglows identified (Cenko et al. 2013, 2015).

The signature of an off-axis afterglow is the presence of a rising then declining source across multiple bands (from the X-ray to the radio) with an apparent power-law spectral slope. Such afterglows are appealing because they are visible on timescales of days, weeks, and even months after a given trigger and can, therefore, be searched for at relative leisure. They are less appealing because they are faint. In particular, although they rise much later, they never exceed the luminosity that would have been seen for an on-axis event at the same epoch. Hence, they are likely to be challenging to detect, even at moderate distances in the absence of preexisting accurate positions.

9.6 Electromagnetic Searches in Black Hole–Black Hole Mergers

The discovery of GW 150914 immediately led to the start of major campaigns of observations to search for associated electromagnetic transients. It was only detected by two detectors and thus had a large error region of 630 deg^2 (90% credibility) spread over both a northern and a southern region (essentially two regions on the annulus created by the 7 ms time delay between Hanford and Livingston; Abbott et al. 2016c). Indeed, a substantial part of the northern portion of this was close to the Sun and challenging to observe (see Figure 9.6).

Nonetheless, multiwavelength observations were obtained by no fewer than 63 separate teams across the electromagnetic spectrum (Abbott et al. 2016c). The first plausible report of a counterpart arose from the analysis of data collected by the *Fermi* GBM, which reported a weak, short-duration GRB beginning approximately 0.4 s after the merger (Figure 9.7). The event was not formally triggered, but an analysis of the likely false-alarm probability led to the conclusion that the chance probability was small and the coincidence therefore significant at the 3σ level (Connaughton et al. 2016; 2018). However, this conclusion was not universally supported. For example, the burst was not detected by the *INTEGRAL* satellite (Savchenko et al. 2016), and other analyses of the same data led to the suggestion that it was, in fact, consistent with a background fluctuation (Greiner et al. 2016). The identification of a GRB from a black hole–black hole merger would be of great interest, as it would provide fundamental insight into material that must surround (and escape) from the merger itself. However, at present, the reality of the detection remains controversial, and it is likely that only further detections (or a large number of nondetections) will resolve the tension.

While the gamma-ray observations could see the whole sky and so potentially claim an association based on time coincidence alone, the same was not true at other wavelengths, and major campaigns were undertaken in an attempt to identify the source. At the time of these observations, the details of the source properties (e.g., type of merger, distance, etc.) were not available to most observers, so different strategies were undertaken. Some searches with wide-field instrumentation attempted to map large portions of the error box and perform systematic searches for transients within it (Smartt et al. 2016). Others noted that the southern portion of the error box overlapped with the Magellanic Clouds, so they focused their efforts on this region in case, for example, a new supernova had appeared, and others attempted to map the most luminous nearby galaxies within the error region (Evans et al. 2016).

No compelling counterpart sources were uncovered. However, many unrelated supernovae were found, all of which appeared to be normal, at large distances, or old. Hence, while the observations were unsuccessful in the discovery of an unambiguous electromagnetic counterpart, they did highlight many of the challenges that are likely to face future searches, especially of large error regions.

Indeed, once the full details of the source were published, it became clear that the distance considered was large, 410_{-180}^{+160} Mpc ($z = 0.09_{-0.04}^{+0.03}$), well beyond the region where complete galaxy catalogs exist. The inability to limit the galaxies surveyed by

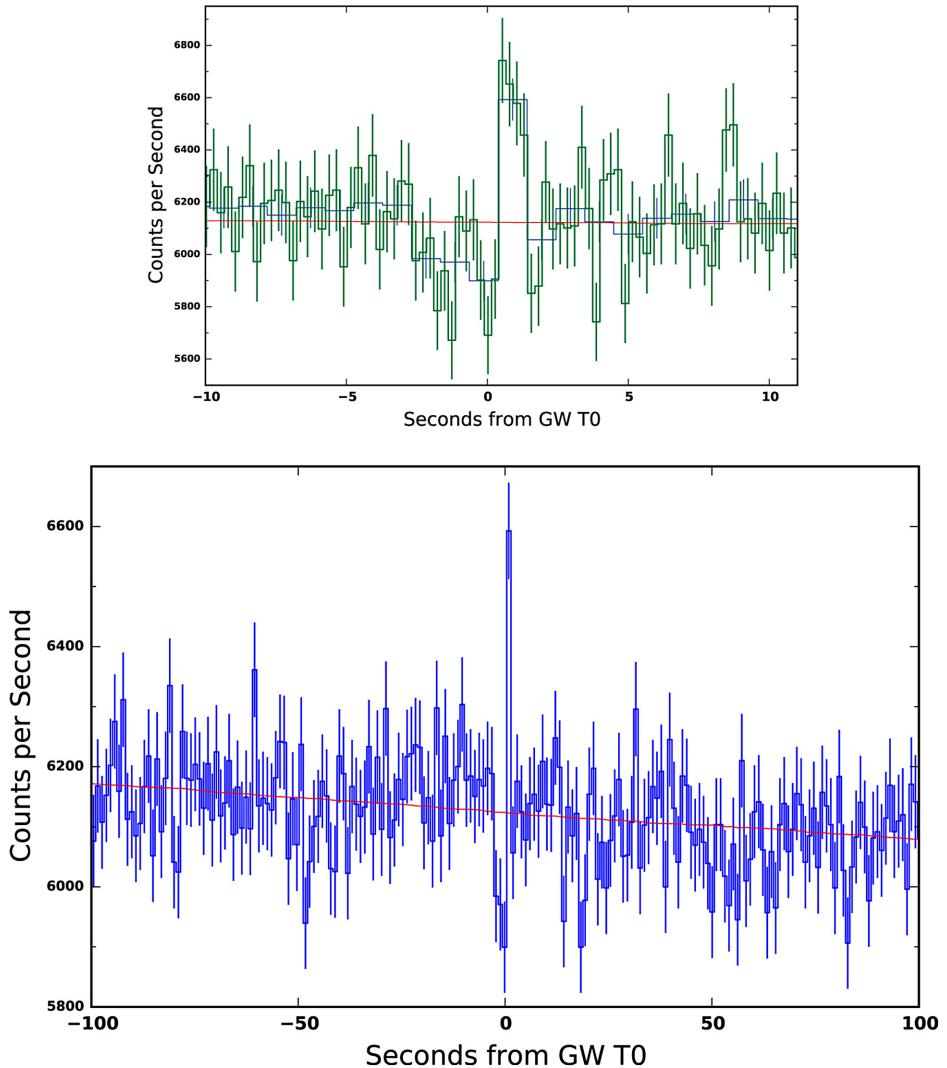


Figure 9.6. Gamma-Ray light curves coincident with the detection of the binary black hole merger GW 150914 from Connaughton et al. (2016). There is an apparent short-duration GRB signal visible in the *Fermi* GBM 0.4 s after the time of the merger. The reality of this signal has been debated (Greiner et al. 2016), although reanalysis of the data yields a false-alarm probability that is still low ($\sim 3\sigma$; Connaughton et al. 2018). © 2016. The American Astronomical Society. All rights reserved.

restricting to those consistent with the gravitational wave-determined distance to the source makes identification more challenging, since almost all detected transients must be considered as possible counterparts in the first instance. Indeed, the strongest route to the removal of such counterparts would likely be a high-cadence survey of the entire visible sky on a nightly basis, so that “old” (i.e., existing before the gravitational-wave trigger) transients could be removed.

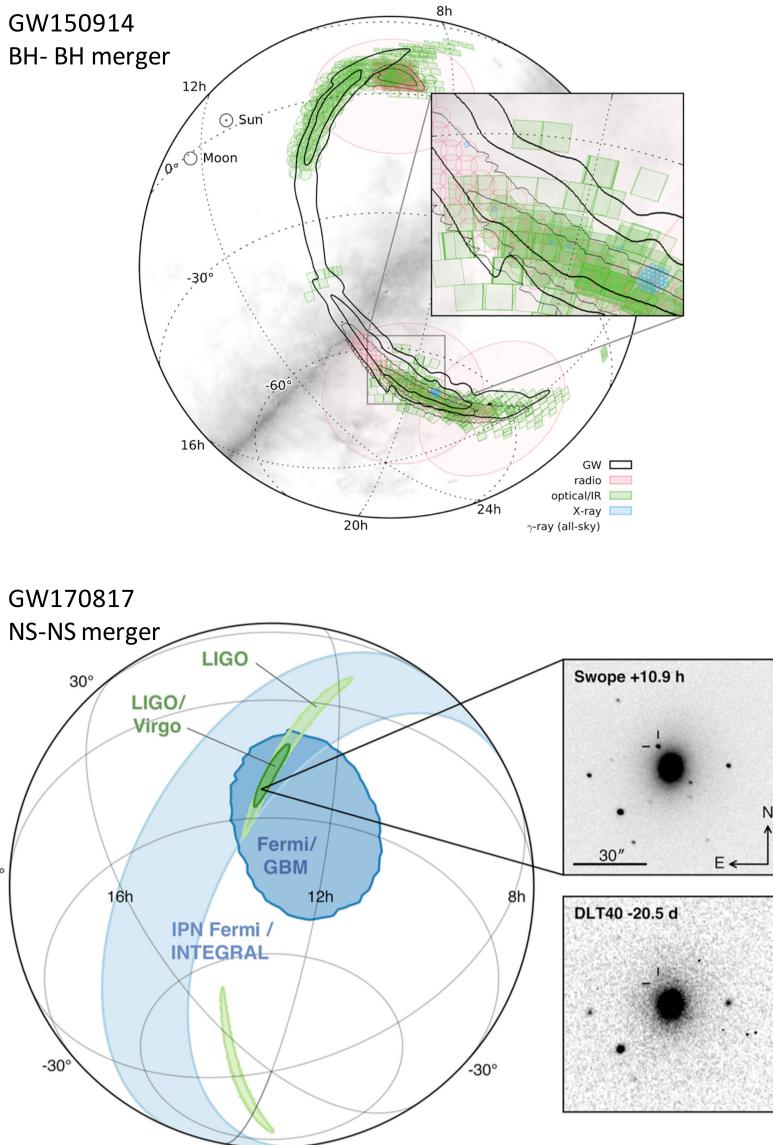


Figure 9.7. Sky localizations of some of the first detected gravitational-wave sources, the black hole–black hole merger GW 150914 (Abbott et al. 2016d; top) and the neutron star–neutron star merger GW 170817 (Abbott et al. 2017d; bottom). The first detection was at high signal-to-noise, but only with two detectors operational. Hence, there is a very large error region. The shaded regions are the coverage of this region in different parts of the electromagnetic spectrum. Remarkably, given the size of the error box, the majority was covered by observations of some kind, although many of these were shallow. For GW 170817 (bottom), similar regions can be seen based on the LIGO detections alone. However, since VIRGO was also operational, it could provide constraints on the source orientation, despite an apparent nondetection of the signal. The resulting error box removes the southern part of the error box and dramatically improves it. In this case, a GRB was also seen, and its location is entirely consistent with the GRB source (as shown by the *Fermi* GBM error box and the annulus created using the time delays from *Fermi* GBM and *INTEGRAL*). © 2017. The American Astronomical Society. All rights reserved. CC.BY.3.0

Similarly, intensive campaigns of observations were launched at each newly identified black hole–black hole merger. No gamma-ray counterparts have been identified in any of these cases, and in the majority, there has been little sign of any unusual transients that could be considered as candidate counterparts for the gravitational-wave source. The one exception to this was the discovery of ATLAS 17aeu (Stalder et al. 2017) in the error box of GW 170104 (Abbott et al. 2017c). This event appeared like a GRB afterglow and decayed rapidly. Such GRB afterglows should be rare in gravitational-wave error boxes; however, in this case, the transient appears to be associated with GRB 170105A, which occurred 19 hr after the gravitational-wave merger and was detected by three satellites within the IPN. It therefore appears unlikely that it is related to GW 170104, unless some very unusual mechanism can be constructed to create a GRB many hours after a black hole–black hole merger or the transient was genuinely related to GW 170104 and not GRB 170105A.

It is important to note that the presence of electromagnetic counterparts to black hole–black hole mergers is unlikely. It is expected that these mergers should be clean, with little baryonic material surrounding them. In this case, electromagnetic counterparts would be unlikely to be detectable with current technology. Even in the scenario that some mass or fossil disk was present around the merger, in most cases, the resulting transients would be very faint, and given the relatively large distances at which the black hole–black hole mergers are found (for example, GW 170104 was at $z \sim 0.2$), their detection would prove very challenging (de Mink & King 2017). It is an important question to ascertain if any emission can be found in the case of black hole–black hole mergers, although it is also increasingly likely that precious observational resources will predominantly be directed at mergers containing neutron stars.

9.7 GW 170817 and GRB 170817A

A breakthrough moment in the search for electromagnetic counterparts to gravitational-wave sources arose on 2017 August 17. At 12:41:04, the LIGO detectors in Hanford triggered on the likely signal of a neutron star binary merger. The source was also visible in the detectors at Livingston, although it overlapped a noise artifact (Abbott et al. 2017e). It was not visible in the VIRGO data, although the strength of the signal in the other two detectors, combined with the noise sensitivity of Virgo at the same time, meant that this nondetection was particularly constraining on the position of the source (Abbott et al. 2017e). The positional error is shown in Figure 9.7 and a timeline of observations in Figure 9.8.

Just 2 s after the detection of the merger (and before its public dissemination), the *Fermi* GBM triggered on the short-duration GRB 170817A. Since short GRBs are rare events, detected by the GBM at a rate of perhaps one per week, the probability of a chance alignment within 2 s of the first binary neutron star merger detected in gravitational waves was small. The error box of the burst was large (15° radius at 50% containment; Goldstein et al. 2017) but vitally overlapped the location derived from the gravitational-wave measurements. The burst was also visible in data taken

by the ESA *INTEGRAL* satellite (Savchenko et al. 2017), providing the opportunity to use time delays to create an annulus and narrow down the resulting error region. This region was also consistent with the location determined from gravitational-wave observations alone. The probability of chance alignment in time and spatially was estimated to be $\sim 5 \times 10^{-8}$, strongly indicative of a connection between the two events.

Motivated by the identification of an associated short GRB, many groups planned observations in the optical and IR, while *Swift* also initiated a program of UV and X-ray observations. The gravitational-wave detection alone had provided an estimate of the distance of the source as 40 Mpc. Unlike the previously detected black hole–black hole mergers, which had distances of hundreds to thousands of Mpc, this distance is sufficiently low that detailed and relatively complete galaxy catalogs exist. Hence, it was possible to determine a relatively short list of plausible host galaxies within the error region (Cook et al. 2017). Indeed, given the shape of the galaxy luminosity function, a significant fraction of the stellar mass (and hence, the likely neutron star binaries) was contained within a rather small number of galaxies. Searches of these could well be effective. This was particularly important because the location of the source was, unfortunately, only $\sim 50^\circ$ from the Sun and thus visible from the best-placed observatories for an hour after sunset.

The first major site where observations were possible was in Chile, where twilight began approximately 11 hr after the trigger, and data were taken from a large number of observatories. The first to observe the galaxy NGC 4993 was the Swope telescope, which observed the galaxy 10.86 hr after the trigger and reported the discovery of a candidate counterpart soon after that (Coulter et al. 2017). NGC 4993 was ranked number 3 (i.e., the third most luminous galaxy) in one list of galaxies within the gravitational-wave error region (Cook et al. 2017) and thus was observed by several groups. Indeed, a total of six groups observed the source location independently, including those using the DLT40 telescope (11.08 hr; Valenti et al. 2017), the VISTA infrared telescope at Paranal (11.24 hr; Tanvir et al. 2017), the MASTER telescope in Argentina (11.31 hr; Lipunov et al. 2017), the Dark Energy Camera on the CTIO 4 m (11.40 hr; Soares-Santos et al. 2017), and the Las Cumbres Observatory (11.57 hr; Arcavi et al. 2017). Evidence suggested that the source was not present in the galaxy in observations taken a few weeks earlier, importantly indicating that it was not an old supernova. The probability of observing a random new supernova within a small sample of galaxies at 40 Mpc is low, making this source, known as AT 2017gfo, a likely counterpart to GW 170817.

Securing the association was not possible with observations on the first night, and the only available spectrum from that epoch showed an apparently featureless blue source (Shappee et al. 2017). Intensive spectroscopic observations were pursued in the nights that followed, and these showed that the source evolved from a hot (~ 8000 K) blackbody into something that was much cooler and, importantly, began showing strong deviations from a purely thermal spectrum, appearing as a series of broad features particularly visible in the infrared (Smartt et al. 2017; Pian et al. 2017; Tanvir et al. 2017). The resulting spectral series from the VLT/X-shooter and a light

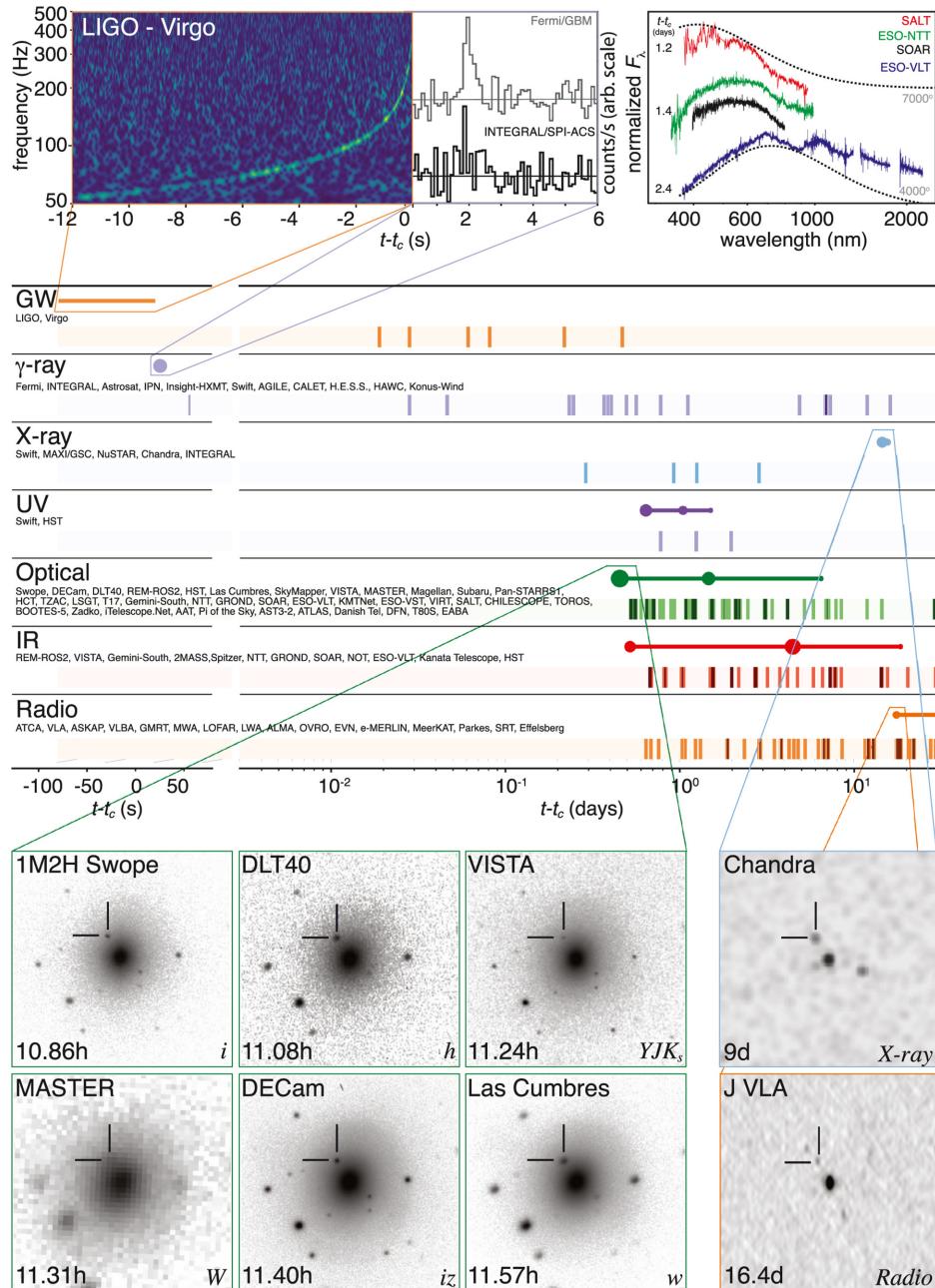


Figure 9.8. Timeline of observations of GW 170817 (from Abbott et al. 2017d) from the detection of gravitational waves prior to the merger, to the identification of the GRB with *Fermi* and *INTEGRAL*, to the location of the source in multiple optical/IR observations 12 hr after the merger. At much later times (several days), the source is also clearly visible to the *Chandra X-ray Observatory* and the Jansky Very Large Array. © 2017. The American Astronomical Society. All rights reserved. CC.BY.3.0

curve (from combined observations; Villar et al. 2017) are shown in Figures 9.9 and 9.10. The strongly chromatic evolution is striking and, importantly, is unprecedented. No other well-studied astrophysical source has been found to show this behavior. With a peak absolute magnitude of $M_R \sim -16$, such events would have been located by preexisting transient surveys should they be common. The unprecedented properties of AT 2017gfo therefore strongly suggested that it was the counterpart of GW 170817.

Indeed, this conclusion is strengthened, since the spectral and photometric evolution is qualitatively very similar to that expected of a kilonova (see Chapter 6), in which the synthesis of heavy, neutron-rich, *r*-process elements proceeds within the ejecta from a neutron star merger. Lighter elements, though still much heavier than iron, are ejected from a disk wind in a predominantly polar direction and create an early, rapidly evolving blue component. Heavier elements are synthesized mainly in the tidal ejecta and have much higher opacities (in particular due to the open *f*-shells in lanthanides), thus creating a source that is very red and peaks on timescales of perhaps $\sim 10\text{--}20$ days.

Outside of the prompt gamma-ray and optical regime, detections of AT 2017gfo proved to be more challenging. Early X-ray observations failed to detect any source (Evans et al. 2017), and radio observations at epochs of a few days also did not recover any source other than the apparently active nucleus of the host galaxy, NGC 4993 (Hallinan et al. 2017). However, observations in these regimes did start to yield detections on timescales of a few days (Troja et al. 2017; Margutti et al. 2017; Hallinan et al. 2017). This suggests that the early optical and IR light was dominated by energy created through the synthesis of heavy elements and was predominantly thermal, but nonthermal emission was becoming visible at later times. The most natural explanation for this is that the X-ray and radio observations arise from a synchrotron component similar (if not identical) to those seen in short-duration GRBs but for which the jet was not originally oriented in our direction. In this case, observers should not initially see the jetted emission but begin to do so at later times as the jet decelerates into the surrounding medium. This would appear to be consistent with the apparent detection of the optical counterpart at late times after the kilonova has faded. The gravitational-wave data prefer an angle of $\sim 30^\circ$ (Abbott et al. 2017e; Mandel 2017) and thus would be consistent with this interpretation.

The off-axis model is also appealing to explain another important question relating to GW 170817/GRB 170817A: why do the apparent energetics of GRB 170817A fall four orders of magnitude below those typical of short GRBs? If the burst were viewed off-axis, it would naturally reduce its energy. However, in this scenario, it is essential to ascertain if an on-axis observer would have seen a powerful short GRB (i.e., if the jet was highly structured so the inferred isotropic energy release was a strong function of the viewing angle) or if the outflow in this case was very different from that in short GRBs, perhaps in the form of some moderately relativistic cocoon (Mooley et al. 2017). Both scenarios have been discussed at length (Lyman et al. 2018; Margutti et al. 2018; Lazzati et al. 2018), although discrimination between them is not possible with the data currently in hand.

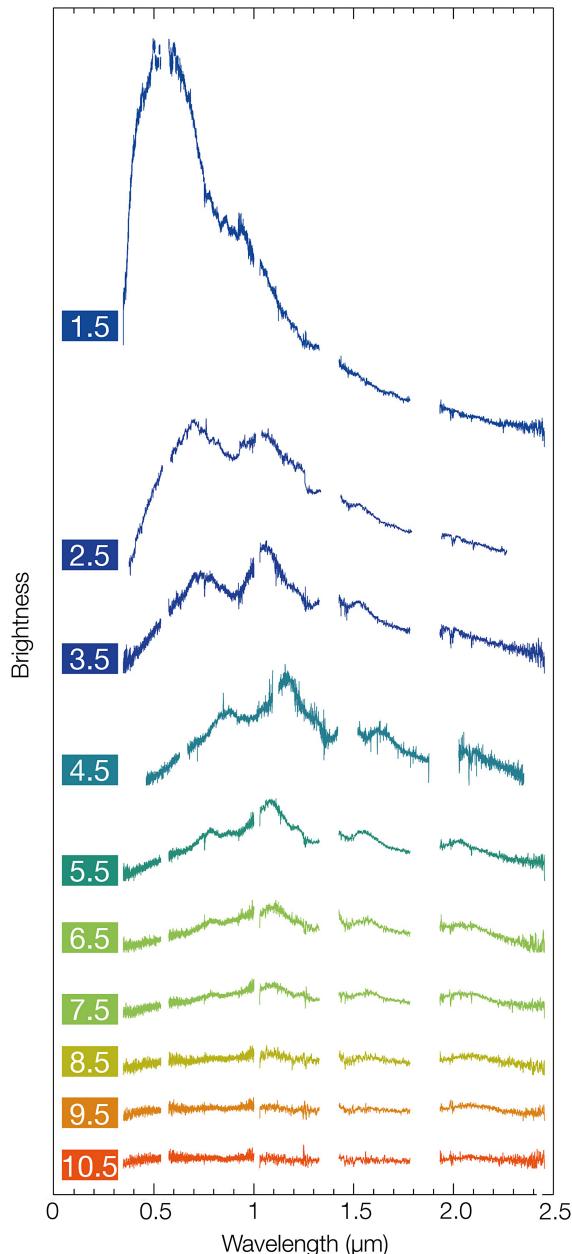


Figure 9.9. Spectral evolution of AT 2017gfo associated with GW 170817 and GRB 170817A. The observations were taken with the X-shooter instrument on the VLT and cover the UV to near-IR in a single snapshot (Smartt et al. 2017; Pian et al. 2017). The strong spectral evolution can be seen with the rapid decay of the optical light and the development of broad features in the IR. The most prominent feature after a few days lies at $\sim 1.1 \mu\text{m}$, approximately the expected location of a previously suggested lanthanide feature (Barnes & Kasen 2013). However, the detailed identification of features has yet to be performed and is challenging because of the lack of laboratory atomic data and high expansion velocities. Image credit: ESO/E. Pian et al./S. Smartt & ePESSTO.

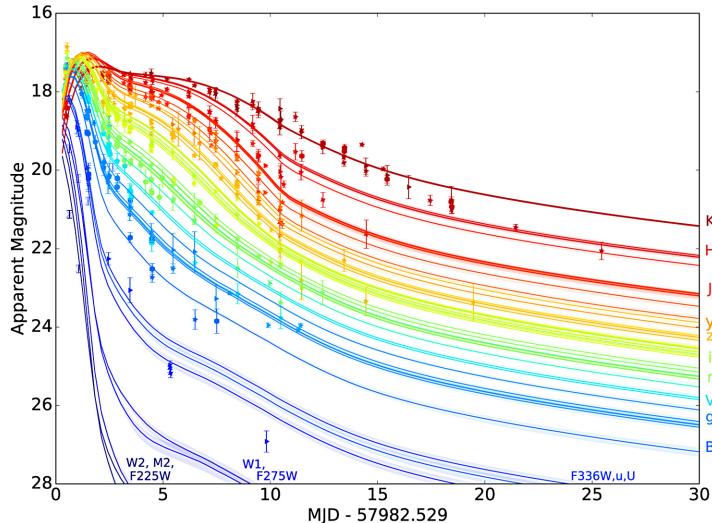


Figure 9.10. Combined optical and IR light curve of AT 2017gfo over the first month from detection. Data are taken from Andreoni et al. (2017), Arcavi et al. (2017), Coulter et al. (2017), Cowperthwaite et al. (2017), Díaz et al. (2017), Drout et al. (2017), Evans et al. (2017), Kasliwal et al. (2017), Lipunov et al. (2017), Pian et al. (2017), Pozanenko et al. (2017), Shappee et al. (2017), Smartt et al. (2017), Tanvir et al. (2017), Troja et al. (2017), Utsumi et al. (2017), and Valenti et al. (2017), as presented in Villar et al. (2017). As with the (lower-cadence) spectroscopy, the rapidly evolving optical source is clearly visible, while the IR begins to dominate at later times, peaking 5–10 days after the merger. © 2017. The American Astronomical Society. All rights reserved.

Having ascertained that GW 170817 is associated with GRB 170817A and AT 2017gfo, it is possible to address the questions that are only accessible via multi-messenger observations. Initially, observations of GW 170817 answer three central questions in contemporary astrophysics. They show that electromagnetic emission can be found in coincidence with a gravitational-wave signal, that short GRBs can arise from the merger of two neutron stars, and that these mergers are a site in which *r*-process nucleosynthesis can occur.

However, they also have a much broader resonance that spans fundamental physics and cosmology. For example, the codetection of gravitational waves and electromagnetic light enables constraints to be placed on various theories of quantum gravity. In particular, in some models, small differences exist between the speed of light and the speed of gravitational waves. These small differences would be amplified by the path length over which the light/gravitational waves have traveled (130 million lt-yr in this case), but none is seen restricting the difference between the speed of light and gravitational waves to less than $\sim 10^{-16} c$ (Abbott et al. 2017b).

Finally, gravitational-wave measurements offer an entirely new route to undertaking cosmological measures, in particular of the Hubble constant, H_0 . Distance measurements are possible directly from the gravitational waves, since the masses of the merging objects (or, more formally, their chirp mass) are accurately constrained

by the waveforms seen during the merger. The masses directly inform the amplitude of the resulting gravitational-wave strain; thus, with knowledge of the mass involved, it is possible to measure the distance directly. This is a so-called standard siren and is analogous to the standard candles that are used very successfully in optical and IR astronomy. The Hubble constant is the constant of proportionality that links the distance of an object with its velocity,⁷ so that $v = H_0 D$. If the distance can be determined through gravitational-wave observations and the velocity via redshift and electromagnetic observations, it is possible to directly measure the Hubble constant.

Measurements of H_0 are still of significant value. As precision has improved in recent years, it has become apparent that the value of H_0 as measured through observations of the CMB ($H_0 = 67.8 \pm 0.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$) based on the recent observations from the *Planck* satellite (Planck Collaboration et al. 2016) are in conflict with those determined from Cepheid variables ($H_0 = 73.48 \pm 1.66 \text{ km s}^{-1} \text{ Mpc}^{-1}$; Riess et al. 2018) at an uncomfortable 3.7σ (Riess et al. 2018). It is plausible that somehow unaccounted-for systematic errors in one or both techniques are leading to this discrepancy, in which case standard cosmological models can still hold. However, if the discrepancy remains, it may be suggestive of new physics that needs to be included, such as additional strains of neutrino or a reevaluation of the so-called cosmological principle, which states that the universe is isotropic and homogeneous on large scales. Indeed, in this latter case, it may be that the local universe where measurements of H_0 are possible with Cepheids is apparently different from the large scales probed with the CMB.

Gravitational waves can resolve this concern because they provide a technique for measuring H_0 that is entirely independent of the other two methods. The first opportunity to make such a measurement is provided by GW 170817 and enables the determination of $H_0 = 70^{+12}_{-8} \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Abbott et al. 2017a). This is still poor precision at present but can be improved with future observations of a larger sample of events. In particular, at 40 Mpc (the distance of GW 170817), peculiar velocities are still particularly important and can bias the results. A large sample would average over such issues, while moving to large distances would increase the expected velocity in the Hubble flow and reduce the importance of the peculiar motion of individual galaxies.

9.8 Gravitational Wave–Electromagnetic Detectors: Questions for the Future

While observations of GW 170817 across the electromagnetic spectrum have provided vital new insight into various astrophysical issues, they have also highlighted further questions. At present, these cannot be addressed with the data in

⁷Formally, this is only true in the local universe before the effects of dark energy become apparent, since the direct proportionality breaks down when this is considered. For distances of <100 Mpc, such an effect is almost always small compared to the errors on other associated measurements, such as the distance.

hand but should be accessible based on additional in-depth study. Of particular importance are the following.

- *What is the mass spectrum of elements ejected in compact object mergers?*

While it is clear that heavy elements, probably up to and including lanthanides, were synthesized in GW 170817, it is less clear exactly what the mass spectrum was. Observations within the solar system provide excellent measurements of the primeval *r*-process abundances, while more limited information is available for other stars via high-resolution stellar spectroscopy. It is therefore relevant to ask if the distribution of atomic masses measured from these observations matches the distribution of ejecta masses from compact object mergers.

This may also relate to the question of the diversity of element production as a function of merger type and viewer angle. For example, if lighter elements are primarily ejected in the form of a disk wind, this may mean that they are preferentially identified in a face-on event and may not be present in an edge-on system that may be dominated by lanthanide production and show no blue component. Furthermore, the details of the total mass ejected are plausibly related to the mass ratio of the merger, with progressively more unequal mass ratios potentially yielding a greater degree of tidal disruption of the less massive (and hence larger) component during the merging process. Although the total mass involved in GW 170817 was very close to $2.8 M_{\odot}$ (depending on the priors used), very close to twice the canonical $1.4 M_{\odot}$ seen in many neutron stars, it is possible that this mass was in fact unequal, with high spin priors enabling mass ratios as small as 0.4 (Abbott et al. 2017e). This may explain why the ejected mass needed to power the luminosity seen in GW 170817 was larger than normally expected.

- *Can compact object mergers account for all heavy-element enrichment in the Universe?* The mass spectrum of the ejecta is closely linked to the broader question of the origin of the *r*-process elements observed in the Universe. There is evidence in the fossil record that favors an origin of the *r*-process elements in rare events that eject significant mass. In particular, radioactive plutonium on the seafloor is out of the equilibrium that would be expected if it were continuously replenished by regular supernovae (Wallner et al. 2015). Similar conclusions are drawn based on studies of europium and strontium in metal stars in the Milky Way (Macias & Ramirez-Ruiz 2016).

The mass release from GW 170817 appears to be large, perhaps $\sim 0.05 M_{\odot}$, and for reasonable astrophysical rates would comfortably be the dominant source of *r*-process elements, assuming the yield per events is similar (Kasen et al. 2017). However, there are still major uncertainties about the per-event yield and rates. Only with the observation of a meaningful sample of events that are representative of the underlying population regarding mass ratio, mass ejection, orientation, etc. will it be possible to conclusively determine the contribution of compact object mergers to the total heavy-element budget in the universe.

- *What do the merger locations tell us about the evolution creating the binary?*

There are many routes to the creation of compact object binaries both in the field (Belczynski et al. 2006) and dynamically (Grindlay et al. 2006). The prevalence of each route remains highly uncertain and is difficult to ascertain from observations of the galactic population. For example, it is possible that a significant fraction of created binaries have short merger times. However, these events are selected against in galactic studies because they have already merged. Similarly, the level of kicks imparted to neutron stars on formation and how this varies with supernova details is an uncertain question (Arzoumanian et al. 2002). Double neutron star binaries are detected by radio emission, normally from the recycled pulsar, but this may not always be visible, and radio surveys may miss events well away from the galactic plane.

The locations of merging neutron stars at merger encode some of this information (see Chapter 6 for information on how this can be applied to short GRBs). Once created, the binaries evolve within the galactic potential until they ultimately merge. For GW 170817, the population in the host galaxy is predominantly billions of years old (Levan et al. 2017), suggesting a very long merger time, but short merger times would enable star-forming regions close to the merger to be identified. Indeed, a population of events associated with star-forming regions may be some of the best evidence for a short delay time population. Finally, unlike short GRBs, which are typically at significant cosmological distances, gravitational-wave sources containing a neutron star are likely to lie at $z < 0.1$. Therefore, it is possible as the source fades to directly identify any underlying globular cluster. Since globular clusters contain only a tiny fraction of the total stellar mass of a galaxy system (<1%), any neutron star binaries located within them are likely to be created dynamically.

- *Can electromagnetic counterparts be found for neutron star–black hole mergers? What about black hole–black hole mergers?* To date, the only electromagnetic counterpart found for a gravitational-wave source is for a system containing two neutron stars. There have been no unambiguously successful multiwavelength searches in the case of black hole–black hole mergers, and any claimed associations remain highly controversial. There is significant interest in the third, hybrid type of merger, that of a neutron star with a black hole. No examples of black hole–neutron star binaries are known within the Milky Way or elsewhere, so their rates of creation are highly uncertain. They are likely to be rarer than neutron star–neutron star binaries, but because of the larger mass due to the presence of a black hole, they can be viewed to larger distances. In principle, this unequal mass ratio could also result in the ejection of a larger quantity of mass during the merging process as the neutron star is “ground down” on successive periastron passages during the merger (Davies et al. 2005). This may yield a bright and detectable optical counterpart, and these sources could also be important in heavy-element production.

9.8.1 Future Prospects for Gravitational-wave Detectors

Our ability to answer the above questions depends centrally on the population of binaries uncovered by gravitational-wave detectors in the next few years. The LIGO and VIRGO detectors are currently undergoing a series of upgrades. The first of these became operational in 2015,⁸ and they will continue until the early 2020s. After each series of upgrades, there is an associated science run of varying direction, and two of these (O1 and O2) have been completed at the time of writing. Each upgrade improves the strain sensitivities of the detectors and hence their horizons and astrophysical reach. In the final sensitivity, the horizon (averaged over sky location and inclination) should enable the detection of binary neutron stars to a distance of 400 Mpc (Abadie et al. 2010). Given the scaling of horizon distance directly with sensitivity, the increase in the expected rates is impressive, and while these remain uncertain, it is likely that when fully sensitive, the detection of black hole–black hole and neutron star–neutron star binaries will be routine and happening at a rate of several per month. There are additional upgrades that can be performed within the LIGO facilities. The first (nearest timescale) of these is known as A+, while the latter is often referred to as Voyager and reflects a longer-term goal. These should ultimately create a facility with a reach >1100 Mpc and far higher astrophysical rates.⁹

In addition to improvements to the existing detectors, there are also plans for further detectors. The KAGRA detector is now well advanced in Japan, and there are advanced plans for a LIGO detector in India. The addition of further detectors is particularly valuable for multimessenger studies, since it dramatically improves the size of the uncertainty regions that must be searched. For example, the fully functioning Voyager network might be expected to deliver 100 binary neutron star inspiral events per year with error boxes smaller than 1 deg² (Mills et al. 2018).

In the longer term, there are also plans for so-called third-generation (3G) detectors. The European Einstein Telescope concept is relatively well advanced. If built, it will be an underground facility with an arm length of 10 km. However, in contrast to LIGO/VIRGO, the layout of the detectors would not be with orthogonal arms but rather with three arms at 60° separations to form a triangular detector. This will offer a further order-of-magnitude improvement in the sensitivity relative to Advanced LIGO, and thanks to the linear scaling of horizon with sensitivity, it will push the horizon for binary neutron star mergers out to $z \sim 3$, while it should effectively observe almost every black hole merger in the visible universe.

On a similar timescale, there are also advanced plans to launch space-based gravitational-wave detectors. These detectors avoid the seismic noise issues that plague ground-based observatories and thus can operate in much lower-frequency regimes. The ESA mission *LISA* is scheduled for launch in the early 2030s, although its pathfinder has recently demonstrated that much of the necessary technology is

⁸Formally, this upgrade became active just *after* the first detection of a binary black hole merger, since the detection of GW 150914 took place during an engineering run.

⁹<https://dcc.ligo.org/public/0125/T1600119/004/wp2016.pdf>

now ready. The design of *LISA* is to use three satellites, each sending a beam to both of the other two while flying in a triangular formation, providing three baselines within a single observatory. They will be sensitive to strains as low as $\Delta L/L = 10^{-23}$, thanks mostly to their extreme separation of approximately 3 million km.

Low-frequency gravitational-wave detectors can identify sources with much wider (longer-period) orbits at times potentially long before they merge. Because of this, the observatory can perform triangulation by measuring the variation of the signal strength as the observatory moves through space. Some of these sources will be persistent, for example, double white dwarf binaries within the Milky Way that have very long merger times (from their current separation). Other sources are transient but will be in the band for much longer than the binaries will be in the LIGO band. For example, a binary like GW 150914 would enter the *LISA* frequency range months before the merger. In principle, a system like this one could then be precisely triangulated well before it entered the LIGO frequency range and could be observed in the process of merging with a range of multiwavelength telescopes to search for any prompt electromagnetic component. Another prime source for *LISA* is the merger of two supermassive black holes ($>10^6 M_\odot$). Such mergers are a vital component of hierarchical galaxy formation models, since when galaxies merge, their black holes should coalesce due to dynamical friction, forcing them to sink through the merged galaxy until the point that gravitational waves take over and drive the merger. Again, *LISA* should observe these signals well before the merger, enabling excellent positions and the possibility of coordinated multimesenger observations.

9.9 Neutrino Emission

The story of the search for astrophysical neutrinos in some ways mirrors that of the search for gravitational waves. Although there was an impressive serendipitous detection of 11 neutrinos from SN 1987A in the Japanese super-kamiokande detector (Hirata et al. 1987), no other detection was forthcoming from these detectors. Indeed, they were tuned for neutrinos of relatively low energy, with a particular focus on solving the solar neutrino problem.¹⁰

Recent very large volume neutrino detectors, such as the IceCube experiment in Antarctica (Karle et al. 2003) or the Antares experiment in the Mediterranean Sea, operate at very different energy ranges and remarkable sensitivities. These detectors use the Earth as a shield that blocks essentially all incoming particles except for neutrinos (that have extremely small cross sections). Hence, rather than looking for particles incoming from above the detector, they are searching for neutrinos coming through the Earth with an upward trajectory in the detectors. They do this with a

¹⁰The solar neutrino problem refers to the fact that the measured solar flux of neutrinos appears to be a factor of 3 lower than predicted based on the nuclear fusion rates required to provide the Sun with its measured luminosity. It was a major problem in astrophysics for many years and suggested major inadequacies in our understanding of either nuclear or particle physics. It now seems likely that neutrino oscillations, in which electron neutrinos (ν_e) can oscillate into those associated with muons (ν_μ) or τ particles (ν_τ), substantially solve this problem.

focus on muon neutrinos that create charged muons on interactions. These, in turn, travel faster than the speed of light in the detector (which is either water or ice) and so emit Cherenkov light. In principle, measuring the properties of these interactions makes it possible to both re-create the track of the neutrino (and hence its direction) and measure its energy.

These detectors are sensitive to very high energy neutrinos, with energies of TeV or higher. Although it is difficult to isolate pure astrophysical signals from those induced from, e.g., cosmic-ray showers in the atmosphere on the opposite side of the planet, there is now good evidence that astrophysical signals are being detected (IceCube Collaboration 2013; Aartsen et al. 2014), since the rate is well above that expected for atmospheric events (see Figure 9.11).

Unfortunately, to date, these are not multimessenger signals, as it has not been possible to identify electromagnetic (or gravitational-wave) counterparts for any individual neutrino source. Indeed, although intensive observations are now regularly undertaken for individual neutrino triggers (Evans et al. 2015), these have yet to be successful.

However, GRBs are prime sources for the creation of such neutrinos (and indeed of ultra-high-energy cosmic rays; see below) because their shocks are likely highly effective particle accelerators (Waxman & Bahcall 1997). In particular, within a GRB shock, it is possible to create highly energetic pions through interactions between protons and photons in the so-called $p\gamma$ interaction (Biehl et al. 2017),

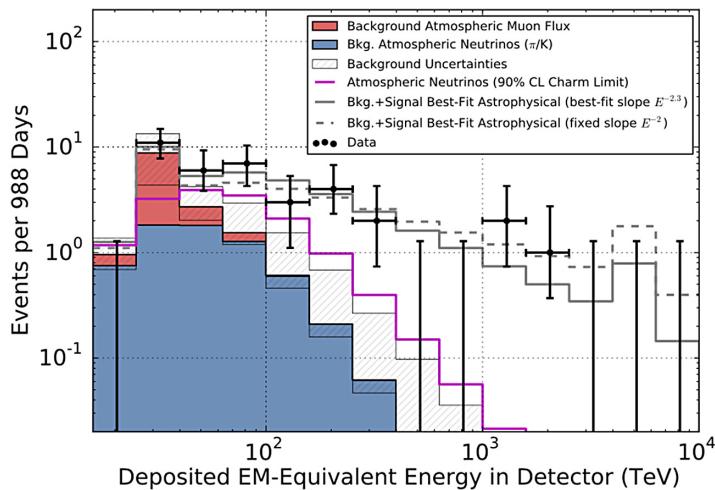


Figure 9.11. Number of events as a function of energy deposited in the IceCube detectors (from Aartsen et al. 2014). Estimates of the backgrounds are shown in the shaded blue and red regions, with uncertainties encompassed by the magenta line. The observed data are shown with error bars and suggest a significant population of very high energy neutrinos that are astrophysical in origin. Since the background here is extremely low, these events have a high probability of association with astrophysical sources and can be pinpointed from the detectors with arcminute precision. Reprinted figure with permission from Aartsen et al. © (2014) by the American Physical Society.

which entails the creation of a so-called Δ resonance (three quarks of the same type) and, subsequently, a charged pion,

$$p + \gamma \rightarrow \Delta^+ \rightarrow \pi^+ + n, \quad (9.4)$$

in one-third of cases (the other two-thirds produce a neutral pion and a proton). The pion subsequently decays to create a muon and muon neutrino,

$$\pi^+ \rightarrow \mu^+ + \nu_\mu. \quad (9.5)$$

The muon then also decays, creating two further neutrinos,

$$\mu^+ \rightarrow e^+ + \nu_e + \overline{\nu}_\mu. \quad (9.6)$$

Hence, these $p\gamma$ interactions have the potential to create a significant neutrino flux from a GRB. The GRBs are promising sites for this process because the conditions for $p\gamma$ interactions can be met. In particular, the protons must be accelerated such that (Zhang & Kumar 2013)

$$E_p E_\gamma \sim \frac{m_\Delta^2 - m_p^2}{2} \left(\frac{\Gamma}{1+z} \right)^2 = 0.147 \text{ GeV}^2 \left(\frac{\Gamma}{1+z} \right)^2, \quad (9.7)$$

where E_p and E_γ are the respective proton and photon energies, m_Δ and m_p are the masses of the Δ resonance and the proton, z is the redshift, and Γ is the bulk Lorentz factor of the associated relativistic outflow. Within a GRB jet, conditions for this reaction are promising, since the photon energy is characteristically at hundreds of keV, the Lorentz factor is typically several hundred, and the resulting neutrino energies will lie in the region of 10^{14} eV ($\sim 0.05E_p$; Waxman & Bahcall 1997). If a significant fraction of the fireball energy is converted into neutrinos then we might expect to observe a similar total energy in neutrinos and photons. In the case of equal total energy release the ratio of the number of photons to neutrinos is equal to the ratio of their individual energies (e.g., the numbers are given by $N_\nu \sim E_\nu/10^{14}$ eV for a neutrino and $N_\gamma = E_\gamma/10^6$ eV for a photon, so that if the total energies are comparable, $E_\gamma \sim E_\nu$, then $N_\gamma/N_\nu \sim 10^8$). At first sight, it would appear that rather few neutrinos may be detected given the differing numbers and the small cross-section for neutrino interactions. However, neutrino detectors are much larger than gamma-ray detectors and may have effective areas of km^2 compared to $< 1 \text{ m}^2$ for the *Swift* BAT. Indeed, Waxman & Bachall (1997) estimated that a burst at 100 Mpc producing 4×10^{50} erg in neutrinos with an energy of 10^{14} eV could be detected with a rate of 3 km^{-2} muons.

These rates are small, and indeed most GRBs observed lie well beyond 100 Mpc (and those at such small distances are typically underenergetic compared to the more distant bursts). Hence, it is necessary to stack data taken from neutrino detectors coincident with many GRBs to obtain stronger constraints. However, since the number of well-localized GRBs is large, this is now possible. Interestingly, the resulting limits lie well below the models, by almost a factor of 4 (Abbasi et al. 2012). This suggests that GRBs are not the sole accelerators of the highest-energy charged particles and that neutrino-producing mechanisms are not, for some as-yet-unknown reason, effective in their ejecta.

9.10 Ultra-high-energy Cosmic Rays

The origin of the most energetic cosmic rays, so-called ultra-high-energy cosmic rays (UHECRs), those with incident energies in excess of 10^{18} eV (Hillas 1984), is tightly linked with the identification of astrophysical neutrinos. The physical processes that produce the neutrinos are themselves the hallmarks of those processes that are likely responsible for the acceleration of cosmic rays.

However, cosmic rays are charged particles, are massive, and, while relativistic, do not travel at the speed of light. They are also subject to deviation caused by the effects of magnetic fields in the interstellar medium. Hence, directional or timing information provides no information about the source of the UHECRs that we observe, whereas the detections of neutrinos should be coincident in both time and space.

Nonetheless, the UHECRs have the advantage that they are readily observable via the atmospheric showers that they induce, and this means that their astrophysical rate is well constrained. In particular, this is not just a total rate, based on the detection of some number of rays, but a volumetric rate, since the UHECRs cannot travel an arbitrary distance because they eventually interact with photons in free space (in the form of either the CMB or the extragalactic background light). Indeed, these reactions are essentially the same as those outlined for the production of pions and neutrinos in GRBs above, i.e.,

$$\gamma_{\text{CMB}} + p \rightarrow \Delta^+ \rightarrow p + \pi^0 \quad (9.8)$$

and

$$\gamma_{\text{CMB}} + p \rightarrow \Delta^+ \rightarrow n + \pi^+, \quad (9.9)$$

with the crucial difference that the photon energies are much lower (eV rather than hundreds of keV).

This scattering process effectively removes UHECRs (those with energies $> 5 \times 10^{19}$ eV) that arise from distances beyond a few hundred Mpc, thus implying that they must arise from relatively local sources, and in turn infers the rate of production from within these sources. This limit, known as the Greisen–Zatsepin–Kuzmin limit, or more often the GZK horizon (Abbasi et al. 2008), places strong constraints on the total budget.

These UHECRs may then arise from either persistent sources, such as active galactic nuclei (AGNs), or transient phenomena. It is apparent that AGNs are not sufficiently numerous within ~ 100 Mpc to account for the observed UHECR flux, so transient phenomena must be considered. It was long thought that GRBs, and perhaps particularly the much more common but lower-energy and low-luminosity GRBs, were a promising location for this cosmic-ray acceleration. However, the nondetection of neutrinos from these sources (Abbasi et al. 2012) casts doubt on this scenario, unless many low-luminosity GRBs are currently missed with the GZK horizon. Alternatively, other transient sources—for example, tidal disruption events with relativistic ejecta (Levan et al. 2011; Bloom et al. 2017)—could also be promising locations. To date, in the absence of clear detections of astrophysical

neutrinos from any given class of event, the true contribution of different systems to the UHECR budget remains uncertain.

9.11 Summary

The promise of multimessenger astronomy has long been recognized, but its realization has taken many years of consistent effort to achieve. Even now, when both gravitational wave-electromagnetic and ν -electromagnetic observations have been made, the events for which both sets of information are available are singular (SN 1987A and GW 170817).

Nonphotonic detectors are now finally reaching the sensitivity where the detection of astrophysical neutrinos and gravitational waves is becoming routine. The challenge now is to launch sufficiently comprehensive multiwavelength observing campaigns that the identification of electromagnetic counterparts becomes similarly standard. Indeed, ultimately, one might hope to observe neutrinos, gravitational waves, and electromagnetic light all from a single source, although this moment may be some time away.

Ultimately, multimessenger astronomy offers the ability to address questions that have been inaccessible to date, and in doing so can provide answers to long-standing controversies in astronomy. However, we are right at the beginning of this era, and there is much work to do before this promise is fully realized.

References

- Aartsen, M. G., Ackermann, M., Adams, J., et al. 2014, [PhRvL](#), **113**, 101101
- Aasi, J., Abadie, J., Abbott, B. P., et al. 2014, [ApJ](#), **785**, 119
- Abadie, J., Abbott, B. P., Abbott, R., et al. 2010, [CQGra](#), **27**, 173001
- Abbasi, R., Abdou, Y., Abu-Zayyad, T., et al. 2012, [Natur](#), **484**, 351
- Abbasi, R. U., Abu-Zayyad, T., Allen, M., et al. 2008, [PhRvL](#), **100**, 101101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, [PhRvX](#), **6**, 041015
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, [PhRvL](#), **116**, 241103
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016c, [ApJL](#), **826**, L13
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016d, [PhRvL](#), **116**, 061102
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016e, [LRR](#), **19**, 1
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, [Natur](#), **551**, 85
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, [ApJL](#), **848**, L13
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, [PhRvL](#), **118**, 221101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017d, [PhRvL](#), **119**, 141101
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017e, [ApJL](#), **848**, L12
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017f, [ApJ](#), **847**, 47
- Andreoni, I., Ackley, K., Cooke, J., et al. 2017, [PASA](#), **34**, e069
- Arcavi, I., Hosseinzadeh, G., Howell, D. A., et al. 2017, [Natur](#), **551**, 64
- Arzoumanian, Z., Chernoff, D. F., & Cordes, J. M. 2002, [ApJ](#), **568**, 289
- Barnes, J., & Kasen, D. 2013, [ApJ](#), **775**, 18
- Belczynski, K., Perna, R., Bulik, T., et al. 2006, [ApJ](#), **648**, 1110
- Berger, E., Fong, W., & Chornock, R. 2013, [ApJL](#), **774**, L23
- Biehl, D., Fedynitch, A., Palladino, A., Weiler, T. J., & Winter, W. 2017, [JCAP](#), **1**, 033

- Bloom, J. S., Giannios, D., Metzger, B. D., et al. 2017, *Sci*, 333, 203
- Cenko, S. B., Kulkarni, S. R., Horesh, A., et al. 2013, *ApJ*, 769, 130
- Cenko, S. B., Urban, A. L., Perley, D. A., et al. 2015, *ApJL*, 803, L24
- Connaughton, V., Burns, E., Goldstein, A., et al. 2018, *ApJL*, 853, L9
- Connaughton, V., Burns, E., Goldstein, A., et al. 2016, *ApJL*, 826, L6
- Cook, D., van Sistine, A., Singer, L., & Kasliwal, M. 2017, GCN Circ., 21519, <https://gcn.gsfc.nasa.gov/gcn3/21519.gcn3>
- Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, *Sci*, 358, 1556
- Cowperthwaite, P. S., Berger, E., Villar, V. A., et al. 2017, *ApJL*, 848, L17
- Davies, M. B., King, A., Rosswog, S., & Wynn, G. 2002, *ApJL*, 579, L63
- Davies, M. B., Levan, A. J., & King, A. R. 2005, *MNRAS*, 356, 54
- de Mink, S. E., & King, A. 2017, *ApJL*, 839, L7
- de Mink, S. E., & Mandel, I. 2016, *MNRAS*, 460, 3545
- Díaz, M. C., Macri, L. M., Garcia Lambas, D., et al. 2017, *ApJL*, 848, L29
- Drout, M. R., Piro, A. L., Shappee, B. J., et al. 2017, *Sci*, 358, 1570
- Eldridge, J. J., & Stanway, E. R. 2016, *MNRAS*, 462, 3302
- Evans, P. A., Cenko, S. B., Kennea, J. A., et al. 2017, *Sci*, 358, 1565
- Evans, P. A., Kennea, J. A., Barthelmy, S. D., et al. 2016, *MNRAS*, 460, L40
- Evans, P. A., Osborne, J. P., Kennea, J. A., et al. 2015, *MNRAS*, 448, 2210
- Fairhurst, S. 2011, *CQGra*, 28, 105021
- Goldstein, A., Veres, P., Burns, E., et al. 2017, *ApJL*, 848, L14
- Greiner, J., Burgess, J. M., Savchenko, V., & Yu, H.-F. 2016, *ApJL*, 827, L38
- Grindlay, J., Portegies Zwart, S., & McMillan, S. 2006, *NatPh*, 2, 116
- Hallinan, G., Corsi, A., Mooley, K. P., et al. 2017, *Sci*, 358, 1579
- Hillas, A. M. 1984, *ARA&A*, 22, 425
- Hirata, K., Kajita, T., Koshiba, M., et al. 1987, *PhRvL*, 58, 1490
- Hulse, R. A., & Taylor, J. H. 1975, *ApJL*, 195, L51
- IceCube Collaboration, 2013, *Sci*, 342, 1242856
- Karle, A., Ahrens, J., Bahcall, J. N., et al. 2003, *NuPhS*, 118, 388
- Kasen, D., Metzger, B., Barnes, J., Quataert, E., & Ramirez-Ruiz, E. 2017, *Natur*, 551, 80
- Kasliwal, M. M., Nakar, E., Singer, L. P., et al. 2017, *Sci*, 358, 1559
- Korol, V., Rossi, E. M., Groot, P. J., et al. 2017, *MNRAS*, 470, 1894
- Kotake, K., Sato, K., & Takahashi, K. 2006, *RPPh*, 69, 971
- Kupfer, T., Korol, V., Shah, S., et al. 2018, *MNRAS*, 480, 302
- Lazzati, D., Perna, R., Morsony, B. J., et al. 2018, *PhRvL*, 120, 241103
- Levan, A., Crowther, P., de Grijs, R., et al. 2016, *SSRv*, 202, 33
- Levan, A. J., Lyman, J. D., Tanvir, N. R., et al. 2017, *ApJL*, 848, L28
- Levan, A. J., Tanvir, N. R., Cenko, S. B., et al. 2011, *Sci*, 333, 199
- Li, L.-X., & Paczyński, B. 1998, *ApJL*, 507, L59
- LIGO Scientific Collaboration, Aasi, J., Abbott, B. P., et al. 2015, *CQGra*, 32, 074001
- Lipunov, V. M., Gorbovskoy, E., Kornilov, V. G., et al. 2017, *ApJL*, 850, L1
- Lyman, J. D., Lamb, G. P., Levan, A. J., et al. 2018, *NatAs*, 2, 751
- Macias, P., & Ramirez-Ruiz, E. 2016, *ApJ*, 860, 89
- Mandel, I. 2017, *ApJL*, 853, L12
- Margutti, R., Alexander, K. D., Xie, X., et al. 2018, *ApJ*, 856, L18
- Margutti, R., Berger, E., Fong, W., et al. 2017, *ApJL*, 848, L20

- Metzger, B. D., & Berger, E. 2012, *ApJ*, **746**, 48
 Mills, C., Tiwari, V., & Fairhurst, S. 2018, *PhRvD*, **97**, 104064
 Mooley, K. P., Nakar, E., Hotokezaka, K., et al. 2017, *Natur*, **554**, 207
 Nelemans, G. 2009, *CQGra*, **26**, 094030
 Pian, E., D'Avanzo, P., Benetti, S., et al. 2017, *Natur*, **551**, 67
 Planck Collaboration, Ade, P. A R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
 Pozanenko, A., Barkov, M. V., Minaev, P. Y., et al. 2017, *ApJL*, **852**, L30
 Riess, A. G., Casertano, S., Yuan, W., et al. 2018, *ApJ*, **855**, 136
 Rodriguez, C. L., Chatterjee, S., & Rasio, F. A. 2016, *PhRvD*, **93**, 084029
 Savchenko, V., Ferrigno, C., Kuulkers, E., et al. 2017, *ApJL*, **848**, L15
 Savchenko, V., Ferrigno, C., Mereghetti, S., et al. 2016, *ApJL*, **820**, L36
 Schutz, B. F. 1986, *Natur*, **323**, 310
 Shappee, B. J., Simon, J. D., Drout, M. R., et al. 2017, *Sci*, **358**, 1574
 Smartt, S. J., Chambers, K. C., Smith, K. W., et al. 2016, *MNRAS*, **462**, 4094
 Smartt, S. J., Chen, T.-W., Jerkstrand, A., et al. 2017, *Natur*, **551**, 75
 Soares-Santos, M., Holz, D. E., Annis, J., et al. 2017, *ApJL*, **848**, L16
 Stalder, B., Tonry, J., Smartt, S. J., et al. 2017, *ApJ*, **850**, 149
 Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. 2013, *Natur*, **500**, 547
 Tanvir, N. R., Levan, A. J., González-Fernández, C., et al. 2017, *ApJL*, **848**, L27
 Troja, E., Piro, L., van Eerten, H., et al. 2017, *Natur*, **551**, 71
 Utsumi, Y., Tanaka, M., Tominaga, N., et al. 2017, *PASJ*, **69**, 101
 Valenti, S., Sand, D. J., et al. 2017, *ApJL*, **848**, L24
 Villar, V. A., Guillotin-Pla, J., Berger, E., et al. 2017, *ApJL*, **851**, L21
 Wallner, A., Faestermann, T., Feige, J., et al. 2015, *NatCo*, **6**, 5956
 Waxman, E., & Bahcall, J. 1997, *PhRvL*, **78**, 2292
 Weber, J. 1969, *PhRvL*, **22**, 1320
 Weisberg, J. M., & Taylor, J. H. 2005, in ASP Conf. 328, Binary Radio Pulsars, ed. F. A. Rasio, & I. H. Stairs (San Francisco, CA: ASP) 25
 Zhang, B., & Kumar, P. 2013, *PhRvL*, **110**, 121101

Gamma-Ray Bursts

Andrew Levan

Chapter 10

GRB Astronomy: Summary and Future Outlook

In the 50 yr since their detection, GRBs have established themselves as a unique tool to address many contemporary questions in astrophysics. They provide a route to understanding the evolution and explosion of extremely massive stars; probe black hole formation, particle acceleration, and physics in extreme conditions; have demonstrated their role in ushering in the multimessenger astronomy era; have provided a new route to measuring the redshifts and physical properties of distant galaxies; can be used to probe the buildup of stellar mass and metals across cosmic history; and provide unique constraints on cosmic reionization. This broad resonance across stellar astrophysics, cosmology, and fundamental physics explains the appeal of the field and the level of activity within it. The GRBs continue to play a vital role in our understanding of the universe, and an increasingly large number of researchers use either GRBs themselves or results derived from their study in their work.

10.1 Challenges for the Future

While GRBs have been highly successful in addressing a wealth of astrophysical questions, many fundamental issues remain. There are many questions that still require answers in the GRB field and several high-level goals that have yet to be realized. A selection of these are summarized below.

10.1.1 The Identification of a Well-studied Sample of GRBs at High Redshift

While a number of high-redshift GRBs have been located by *Swift* since it launched, the rate of discovery of these bursts is low, and the total number only small, especially beyond $z \sim 6$, where bursts become most interesting as probes of reionization and the first generation of stars. While breakthroughs in this area can arise from single events—for example, the identification of a GRB from a likely

Pop III star—much of the work relies on the buildup of a sample of GRBs at high- z with good afterglow spectra and host follow-up. This will enable the luminosity function of these bursts to be compared to lower- z examples to search for any evidence of additional populations, for example, from Pop III stars. The afterglow spectroscopy will measure the metallicity of the universe at this epoch, as well as mapping the evolution of the neutral fraction with redshift along multiple lines of sight. Finally, observations of the host galaxies can test whether there is sufficient production of ionizing photons and test the properties of the first galaxies in a manner not available to other selection methods. In practice, the rarity of high- z bursts likely means that this will require a long period of observations with a satellite like *Swift* or the advent of a mission more sensitive for high- z GRB discovery. Ultimately, the capabilities of the next generation of ground-based telescopes, such as the next generation of extremely large (30–40 m aperture) ground-based telescopes, mean that detailed absorption spectroscopy will be possible at far fainter magnitudes than has previously been possible (see Figure 10.1).

10.1.2 The Detection of GRBs from First-generation Stars

Related to the identification of high- z GRBs is the possibility of observing bursts from Pop III events. In principle, these Pop III explosions could be very different to local GRBs, although such a difference is not necessary in all models. If Pop III stars are indeed extremely massive, then the mass reservoir available to accrete onto the nascent black hole could also be very large, potentially yielding extremely energetic

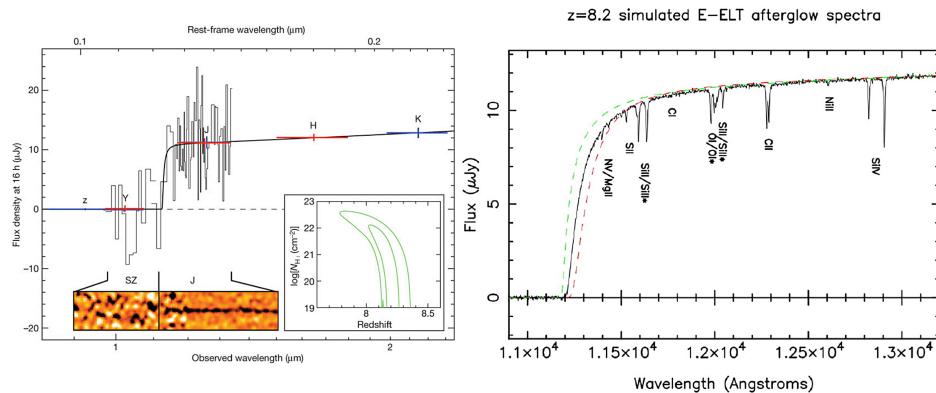


Figure 10.1. Left: observed spectrum of GRB 090423 as observed with the VLT/ISAAC (from Tanvir et al. 2009). Reprinted by permission from [Springer Nature]: [Nature] [International Journal of Science] [Tanvir et al. A γ -ray burst at a redshift of $z \approx 8.2$. Vol. 461, pages 1254–1257] [2009]. Right: simulated spectrum of the same afterglow if observed with an ELT (40 m aperture) for the same exposure time (Amati et al. 2018). The difference in signal-to-noise is striking and largely thanks to the increased mirror area providing a factor of 25 more photons (although it is also helped by advanced adaptive optics that spread the light over fewer spatial pixels). Rather than just providing a redshift measurement, such observations can measure metallicities via narrow metal lines and disentangle the effect of hydrogen absorption from the host galaxy and that from the intergalactic medium. Reprinted from Amati et al. Copyright (2018), with permission from Elsevier.

GRBs. Similarly, since these stars may be very large, the timescale for accretion could also be long, yielding long, highly energetic GRBs. In some ways, the ULGRBs (Levan et al. 2014) appear to meet many of these requirements, and their origin in a Pop III-like star has been postulated (Evans et al. 2014; Piro et al. 2014). However, since these bursts apparently reside at moderate redshifts in galaxies that have at a significant degree of metal enrichment, such an interpretation is unlikely to allow the collapse of stars made from truly pristine (Big Bang levels of enrichment) stars.

10.1.3 The Use of GRBs to Determine Cosmological Parameters

It remains unclear if GRBs can ever be used to reliably determine cosmological parameters in a way that provides meaningfully improved constraints to those possible with other methods. Nonetheless, significant effort continues to be directed at relations that may enable such progress to be made. The ability of GRBs to probe very large luminosity distances means that differences between different cosmologies become larger and thus can be probed with increasingly less precise standard candles. However, as other techniques become more mature and the associated errors on the cosmological parameters become smaller, the possibility that GRBs can make valuable contributions to this field becomes less.

10.1.4 The Detection of More GRBs in Coincidence with Gravitational-wave Triggers and the Addition of Neutrinos

The detection of GRB 170817A in coincidence with GW 170817 was a watershed moment in astronomy, finally demonstrating the origin of short GRBs and showing that multimessenger astronomy between electromagnetic light and gravitational waves was possible. However, while this single event provided a wealth of astrophysical information, there are many more questions that can only be answered by observations of additional systems that will pin down the true rate of the mergers, discern the diversity of associated transients, and enable their contribution to heavy-element production to be calculated. Hence, future multimessenger observations are crucial, both of neutron star–neutron star binaries and of those including a black hole. Indeed, there remains the possibility that black hole mergers could also produce a GRB, and this can be demonstrated or falsified with further observations. It is likely that the joint GRB (or electromagnetic) and gravitational-wave possibilities will be the major focus of GRB research in the coming years.

Beyond the detection of gravitational waves, there is also the possibility of detecting neutrinos in coincidence with GRBs or with GRBs and gravitational waves. Indeed, it has been noted that the current era is really bi-messenger, so adding a third axis of neutrino detection would enable an ever more complete picture to be built.

10.1.5 The Direct Detection of GRB Progenitors

The nature of supernova progenitors was clinched with the direct detection of their progenitors in pre-explosion imaging, now often followed by the confirmation that the progenitor has vanished after the supernova has faded. The detection of GRB

progenitors would be extremely valuable in this regard. However, such work is unlikely to be straightforward. The volumetric rate of GRBs is significantly lower than that for supernovae, even after the significant beaming correction is accounted for. Therefore, the probability of locating a GRB in a local galaxy, where the progenitor is readily identified, is small. However, there are reasons to be optimistic. New surveys are likely to increase the transient rate dramatically; this means that the probability of catching GRBs that are off-axis increases. This can be done at either optical or radio frequencies, where radio observations may provide the ability to detect the source for several years. Coupled with this, deep surveys, such as those conducted by the Large Synoptic Survey Telescope, ESA's *Euclid*, or NASA's *WFIRST*, will provide deep imaging with high resolution across much of the visible sky. This will enable a relatively complete sample of galaxies to be observed out to several tens of kpc, where it may be possible to directly identify luminous progenitors.

It is possible that GRB progenitors may not be especially luminous immediately prior to their collapse, at least at optical wavelengths, so further progress may well be made by narrowband surveys (especially if these can also be wide). Since it is widely believed that the progenitors of long GRBs are Wolf–Rayet stars, they should exhibit strong emission lines, such as He II or C IV. Surveys in these emission lines may therefore be a powerful route to the identification of Wolf–Rayet progenitors, both for long GRBs and hydrogen-deficient supernovae in general.

10.1.6 The Full Characterization of the GRB Blast Wave and Its Geometry

Although it is clear that GRBs are relativistic explosions that are narrowly collimated, the details of this emission remain unclear. Are the jets simple top-hat structures, or do they have more complex angular structure? How do the shocks work? What is the Lorentz factor, and how does it evolve? Is the structure of GRB jets similar in short- and long-duration GRBs? What is GRB beaming and true energetics? All of these questions can, in principle, be answered by in-depth observations of GRBs over the next few years; crucially, this will need to leverage observations across the electromagnetic spectrum, as well as beyond (e.g., via gravitational waves). The identification of populations of bursts that were not initially directed toward us (so-called orphan GRBs) is also central in understanding the details of the blast wave and may be possible with next-generation optical (e.g., LSST) and radio (e.g., SKA) surveys.

10.2 Possibilities for Future GRB Detection Missions

The detection of GRBs, at least to date, has almost exclusively relied on the identification of the prompt gamma-ray signal. This is likely to remain the primary route to detection in the future, although there are possibilities for detecting bursts directly on the ground with the next generation of Cherenkov gamma-ray detectors, using X-rays to detect the burst, or just detecting the afterglow (either on- or off-axis) via wide-field surveys in the X-ray, optical, or radio regimes. Different science goals impose different restrictions on the design of future observatories for GRBs, so

there is no single “next-generation” approach. However, in general, there are several improvements that can be made over currently flying detectors in order to enhance not just the rate of GRBs but the scientific return that each burst enables.

Ultimately, missions are governed by their scientific return and their cost, both of which are the dominant factors in their selection for flight. Such selection is a highly competitive process, in which ideas from across and beyond the astronomical community are solicited by space agencies. These then undergo rigorous review to determine if they are technically plausible and well costed and if the scientific returns are likely. Only after this point can a mission be selected and building commence. It is therefore no surprise that the lead time, even on modest-sized missions, is frequently a decade, while large missions can require several decades of development. Indeed, for comparison, the *James Webb Space Telescope*, currently scheduled for launch at the end of 2018, was first proposed almost 30 yr earlier.¹

The principal concerns that must be considered in the design of a new mission are sky coverage, sensitivity, energy band, and spatial resolution. An ideal detector would see the whole sky to excellent sensitivity across the electromagnetic spectrum and provide subarcsecond positions. Since no technology is known that can do this, compromises must be made in each of these areas, and the decision as to which compromises to make depends on the science that is prioritized by the mission, although choices must often be made based not just on previously observed properties of GRBs but on the theoretically expected signals. Two generic high-profile examples would be missions that were optimized for the detection of GRBs in coincidence with gravitational-wave signals and missions that aimed to detect populations of high-redshift GRBs.

If the aim is to determine the coincidence of short GRBs with gravitational-wave triggers, then sky view is important, while the relative hardness of the short-burst population may argue for a higher-energy threshold. Since coincidence alone can be done based on timing, the positional accuracy may be relatively poor (as it was for the *Fermi* detection of the first gravitational wave–GRB counterpart, GW 170817/GRB 170817A). Alternatively, if the aim of the detection would be to allow multiwavelength observations, then a more precise position would be highly desirable, since the positions available from the gravitational-wave detections alone are seldom expected to be less than tens of square degrees.

If the aim of the observations is to observe GRBs at very high redshift, a rather different approach might be used. First, the impact of redshift is to soften the observed spectrum, so a softer response of the detectors might be needed. Second, it is clear that missions to date have not detected large numbers of high- z GRBs, perhaps because they only observe the brightest examples. It is therefore necessary to consider both the shape of the GRB luminosity function and the rate of GRBs to find a compromise between sensitivity and field of view. More sensitive instruments can go deeper, but often over a smaller region of sky, and these concerns must be weighted against each other to determine the optimum combination.

¹ <https://jwst.stsci.edu/about-jwst/history>

There is a high attrition rate for proposed space missions. The vast majority of those suggested never fly, and many that do undergo many revisions (often de-scopes) from proposal to launch. Therefore, it is very difficult to predict which, if any, of the currently outlined missions will ultimately produce GRBs. However, some of the concepts that are currently under consideration, or in some cases being built, include² the following.

- The *Space Variable Objects Monitor (SVOM)* will likely be the next major GRB mission launched (Paul et al. 2011). It is a primarily Chinese and French collaboration that will build on the success of *Swift* by utilizing a similar concept in which follow-up is conducted rapidly on board the satellite, which will contain gamma-ray, X-ray, and optical detectors. Its main localizing detector, ECLAIRS, will identify approximately 200 GRBs per year, somewhat lower than *Swift*. However, the mission is somewhat better optimized for the detection of high- z GRBs. In particular, the sensitivity of ECLAIRS will extend down to \sim 4 keV, compared to \sim 15 keV for the *Swift* BAT. This will allow for the detection of bursts whose spectra have been significantly softened by redshift. The Microchannel X-ray Telescope (MXT) should identify the X-ray afterglow, while the Visible Telescope (VT) can locate the optical afterglow. In a further effort to increase the haul of high- z bursts, the VT is much larger than the *Swift* UVOT (45 cm versus 30 cm) and is optimized to extend to \sim 950 nm, much redder than the \sim 650 nm cutoff that UVOT observed. Hence, it should be able to identify bursts up to $z \sim 7$. A further addition for *SVOM* is the inclusion as part of the mission of a ground-based component, the Ground Follow-up Telescopes (GFTs) and the Ground Wide Angle Cameras (GWACs). The wide-angle cameras will follow the field of view of the gamma-ray detectors to provide simultaneous optical emission, while the larger (\sim 1 m) GFTs will repoint at the location of any bursts within a few seconds of an alert. This combination should allow the recovery of a far larger fraction of GRB afterglows and enable the identification of new high- z events, in addition to many other GRBs across the redshift range.
- The *Transient High Energy Sources and Early Universe Surveyor (THESEUS*; Amati et al. 2018; see Figure 10.2) is a proposed medium-sized mission for ESA (M5, the fifth of these missions, following *Solar Orbiter*, *Euclid*, *Plato*, and another that has yet to be decided (M4)). It also follows the exceptionally successful *Swift* concept in placing a gamma-ray, X-ray, and optical/IR platform on the same observatory, although it uses somewhat different technology to do this. Crucially, it seeks to directly obtain positions accurate to \sim 1' through the use of wide-field X-ray telescopes that use so-called lobster-eye technology. This enables an X-ray detector with a response extending down to 0.3 keV to observe 1 sr of sky at a given time. This will be paired with a higher-energy detector extending to 20 MeV to provide a more traditional GRB energy range and also an excellent spectral range. This

²This list is not intended to be in any way exhaustive. The list includes proposed and approved missions for which the detection of GRBs will be a major focus.

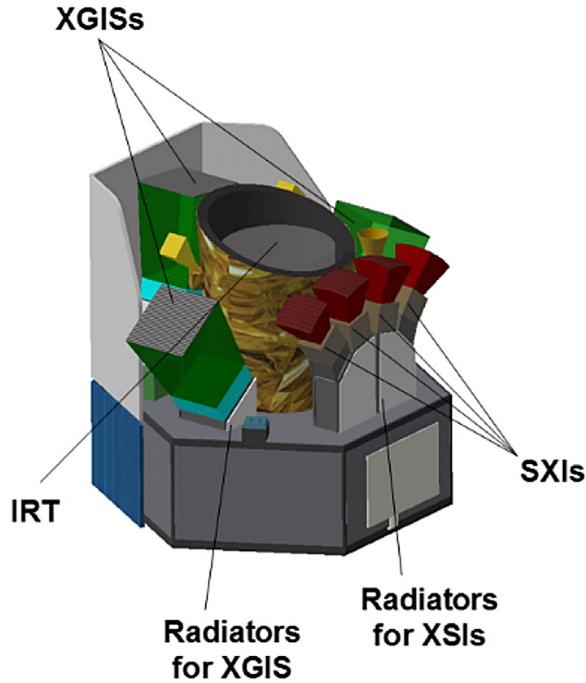


Figure 10.2. Proposed layout of the ESA *THESEUS* mission (Amati et al. 2018). As with *Swift*, it focuses on providing onboard detection and localization. Bursts will be detected by the soft X-ray telescope (0.3–6 keV) using lobster-eye optics, while a hard-coded mask will provide a far wider lever arm than *Swift* (2 keV–20 MeV). The bursts, once discovered, will be further localized by the 0.7 m IR telescope (note that this is of comparable size to the *Spitzer* telescope, one of NASA’s Great Observatories). Reprinted from Amati et al. Copyright (2018), with permission from Elsevier.

will be coupled with a 0.7 m IR-optimized telescope that will crucially also contain spectroscopic follow-up onboard. This route should both enhance the detection rate of high- z GRBs and enable a nearly complete sample of them to be identified on board the satellite, avoiding issues inherent in ground-based follow-up (although the full success of the mission would depend on in-depth follow-up from ground-based facilities, particularly the extremely large telescopes). The soft X-ray response also makes it an extremely powerful tool for identifying other transient sources, many of which do not emit significantly at gamma-ray energies but are copious X-ray emitters. Hence, it should provide by far the most powerful route to date of probing the high-energy transient sky.

- The *Transient Astrophysics Probe (TAP)* is a proposed mission developed following the 2010 US Decadal Review and currently in a mission-concept study phase. It follows a similar model to *THESEUS*, consisting of transient discovery via X-rays and lobster-eye optics (but with a wider 2 sr view), an X-ray telescope with comparable sensitivity to *Chandra*, and a 40 cm IR-optimized telescope. In principle, it should be capable of identifying

a factor of >5 more GRBs than currently observed with *Swift*, as well as finding many more X-ray transients of both the short- and long-lived varieties.

- In addition to “standard” mission concepts, a route to launch that has become increasingly popular in recent years is the idea of cube satellites (CubeSats). These are very small satellites that can be built at low cost and launched alongside other, much heavier payloads at a minimal additional cost. In general, they are only 10–30 cm in size with a mass of a few kg. Individually, therefore, they are not especially capable, but the possibility that many could be launched is appealing, especially if routes can be found to combine their signals to increase sensitivity. There is a particular appeal for gravitational-wave follow-up, since, unlike satellites in low-Earth orbit (such as *Swift* and *Fermi*), an array of CubeSats could provide all-sky coverage. There are various possibilities currently under consideration, including the BurstCube concept.³

10.3 The Crucial Role of Follow-up

An important lesson from observations with *Swift* is the importance of follow-up observations in maximizing the science. The major stride forward made with *Swift* was the ability to conduct its own automated follow-up observations with the X-ray telescope and UV and optical telescope. This enabled precise subarcsecond positions and provided ground-based observers with some basic expectations, all within seconds to minutes following the discovery of a burst. Future missions will often take this one step further and seek to provide onboard information about the GRB redshift, for example. However, for detailed characterization, there will always be the requirement of follow-up from facilities beyond the detecting mission itself. Some of these may be small facilities on the ground—indeed, *SVOM* is notable for its approach to including ground-based follow-up as part of the initial science specification—but success will also depend on leveraging time allocations on large telescopes. As the number of transients (not just GRBs) spirals over the next decade, the identification and prioritization of the most scientifically valuable targets will become a central component of the success of any campaign.

10.4 Summary

The science from studies of GRBs is now important across the bulk of astronomy and even beyond into fundamental physics. The GRBs have been used to provide stringent limits on quantum gravity effects through differing speeds of light at different energies, have been proposed to explain mass-extinction events in the geological history of the Earth, have progenitors for short bursts that may create the majority of heavy elements throughout the universe, are successfully used to probe the physical conditions in the early universe, provide a route to mapping the buildup of stellar mass and metals across cosmic history, provide unique constraints on the processes of particle acceleration and the production of ultra-high-energy cosmic

³ <https://asd.gsfc.nasa.gov/burscube/>

rays and neutrinos, probe the stellar evolution of the most massive stars, and have been demonstrated as route to enabling multimessenger astronomy.

This broad range of uses means GRBs are now very much in the mainstream of astronomy, with an ever-increasing number of astronomers either directly observing or modeling them or using results enabled by them. The advent of multimessenger astronomy suggests that they will remain at the forefront of astrophysical research for some time to come and that many new insights into a very wide range of astrophysical questions can be expected in the coming years.

References

- Amati, L., O'Brien, P., Götz, D., et al. 2018, *AdSpR*, **62**, 191
- Evans, P. A., Willingale, R., Osborne, J. P., et al. 2014, *MNRAS*, **444**, 250
- Levan, A. J., Tanvir, N. R., Starling, R. L. C., et al. 2014, *ApJ*, **781**, 13
- Paul, J., Wei, J., Basa, S., & Zhang, S.-N. 2011, *CRPhy*, **12**, 298
- Piro, L., Troja, E., Gendre, B., et al. 2014, *ApJL*, **790**, L15
- Tanvir, N. R., Fox, D. B., Levan, A. J., et al. 2009, *Natur*, **461**, 1254
- Thatte, N., Tecza, M., Clarke, F., et al. 2010, *Proc. SPIE*, **7735**, 77352I