Gamma-Ray Bursts
The Brightest Explosions in the Universe
Gilbert Vedrenne and Jean-Luc Atteia

Gamma-Ray Bursts
The Brightest Explosions in the Universe

Springer
Published in association with Praxis Publishing
Chichester, UK
Contents

Preface ......................................................... xi
Acknowledgements ........................................... xiii
Foreword ....................................................... xv
List of figures ............................................... xvii
List of tables ............................................... xxiii
Introduction ............................................... xxv

1 The early times ........................................... 1
  1.1 Introduction ........................................... 1
  1.2 Morphology and temporal properties of GRBs ............ 1
     1.2.1 Periods .......................................... 4
     1.2.2 Soft X-ray emission .............................. 4
  1.3 Spectra of GRBs ........................................ 6
     1.3.1 The continuum .................................... 6
     1.3.2 Spectral lines .................................... 12
  1.4 Angular distribution ................................... 16
  1.5 Intensity distribution .................................. 20
  1.6 GRB counterparts ........................................ 29
  1.7 The neutron star paradigm .............................. 33
  1.8 Conclusion ............................................. 37
  1.9 References ............................................. 40

2 The BATSE decade ......................................... 49
  2.1 Morphology and spectra ................................ 49
     2.1.1 Morphology ....................................... 49
## Contents

2.1.2 GRB spectra seen by CGRO-BATSE .......................... 53
2.1.3 Correlations between spectral and temporal properties 57
2.2 Spatial and intensity distributions ............................... 58
  2.2.1 Angular distribution ...................................... 58
  2.2.2 Intensity distribution ..................................... 59
2.3 Possible space distribution of GRB sources ....................... 61
  2.3.1 Single component galactic models ........................ 61
  2.3.2 Two-component models ................................. 64
  2.3.3 Cosmological models .................................... 66
2.4 Searching for cosmological signatures ........................... 66
  2.4.1 Time dilation ........................................... 66
  2.4.2 Spectral softening of dim bursts ........................ 70
  2.4.3 Lensing effects ......................................... 71
2.5 GRB counterparts ................................................ 71
2.6 Beyond the distance problem, some important open issues ... 75
2.7 Conclusion .................................................... 78
2.8 References ................................................... 78

3 BeppoSAX and the afterglow era .................................... 87
  3.1 BeppoSAX: its instruments, its strategy ...................... 88
  3.2 Some GRBs seen in 1997 and their afterglows ................. 89
    3.2.1 GRB 970228, the first X-ray afterglow and optical counterpart 89
    3.2.2 GRB 970508, the first distance and the first radio afterglow .... 94
    3.2.3 The conference on the island of Elba .................. 97
    3.2.4 GRB 971214, a very distant burst ..................... 97
  3.3 Dark GRBs ................................................... 99
  3.4 Jets and collimation ......................................... 102
  3.5 Orphan afterglows ............................................ 107
  3.6 The GRB–supernova connection ............................... 108
  3.7 Emission and absorption features in the X-ray spectra of GRBs and in their X-ray afterglows ................. 113
  3.8 X-ray-rich gamma-ray bursts ................................ 115
  3.9 The connection between the prompt GRB and the afterglow ... 117
    3.10 The $E_{\text{peak}}$–$E_{\text{iso}}$ (Amati) relation ............. 119
    3.11 Conclusion ................................................ 123
    3.12 References ................................................ 125

4 HETE-2 and Swift .................................................. 135
  4.1 The HETE-2 mission .......................................... 135
  4.2 Fast alerts in 2002 ........................................... 138
    4.2.1 GRB 021004: optical follow-up from 193 s to 3 months after the burst ......................... 138
    4.2.2 GRB 021211: a dark GRB caught early ................. 140
4.3 GRB 030329, the GRB and its supernova .......................... 141
4.4 GRB 050709, the first optical afterglow of a short GRB ........ 148
4.5 X-ray flashes, the GRB soft cousins ............................ 151
4.5.1 The discovery of X-ray flashes ............................ 151
4.5.2 HETE-2 confirms that X-ray flashes are soft GRBs ...... 152
4.5.3 The nature of X-ray flashes ............................ 154
4.6 INTEGRAL and the X-ray halos of GRBs ....................... 157
4.7 The Swift mission ............................................. 160
4.7.1 BAT: the Burst Alert Telescope .......................... 162
4.7.2 XRT: the X-Ray Telescope ............................ 163
4.7.3 UVOT: the Ultra-Violet/Optical Telescope .............. 164
4.7.4 The alert system ............................................. 164
4.8 A general analysis of the first GRBs of Swift ............ 165
4.9 The different phases of the X-ray afterflows ............ 167
4.9.1 The canonical early X-ray afterglow light-curve ..... 168
4.9.2 X-ray flares ............................................. 170
4.9.3 Achromatic X-ray breaks? ............................. 177
4.10 Some exciting GRBs ........................................... 177
4.10.1 GRB 060218, the first low-redshift X-ray flash ...... 177
4.10.2 GRB 050904: the most distant GRB .................. 181
4.10.3 GRB 061007: today the brightest GRB ............. 187
4.10.4 GRB 060614 a nearby long-duration GRB without supernova signature ........................................... 189
4.11 Afterglows of short GRBs ................................. 195
4.12 Conclusion .................................................... 197
4.13 References ..................................................... 199

5 Gamma-ray bursts—fireballs and blastwaves ...................... 219
5.1 Fireballs ..................................................... 220
5.1.1 Relativistic expansion and pair opacity .................. 220
5.1.2 Basic fireballs ............................................. 221
5.2 Shocks and photon emission .................................. 224
5.2.1 Relativistic shocks ..................................... 224
5.2.2 Timescales in relativistic flows ......................... 225
5.2.3 Microphysics parameters ..................................... 226
5.2.4 Particle acceleration .................................... 227
5.2.5 Synchrotron emission .................................... 229
5.2.6 Inverse Compton Scattering (ICS) ....................... 233
5.3 The prompt emission ........................................... 234
5.3.1 Internal versus external shocks .......................... 234
5.3.2 Internal shocks ............................................. 235
5.3.3 GRB light-curves and internal shocks .................. 237
5.3.4 Prompt GRB spectra and the synchrotron shock model ........ 239
5.3.5 The ‘line of death’ of synchrotron emission models .. 240
5.3.6 Efficiency of the internal shock model .................. 243
5.3.7 Producing the prompt GRB emission without internal shocks ........................................ 247
5.4 Conclusion ................................................................................................................. 250
5.5 References ................................................................................................................ 252

6 Gamma-ray burst afterglows ...................................................................................... 259
   6.1 The dynamics of the fireball .................................................................................... 259
      6.1.1 Evolution of the fireball .................................................................................... 259
      6.1.2 The temporal evolution of $R$ and $\Gamma$ .......................................................... 263
   6.2 External shocks ....................................................................................................... 266
   6.3 Energy spectrum of the afterglow ............................................................................ 269
   6.4 Afterglow light-curves in the standard model ......................................................... 271
   6.5 The standard afterglow model in detail .................................................................. 275
   6.6 Reverse shock and early optical flashes .................................................................. 276
      6.6.1 Some examples ................................................................................................. 278
      6.6.2 Alternative explanations of early optical flashes .............................................. 281
      6.6.3 Suppressing the reverse shock optical emission? .............................................. 283
   6.7 Non-spherical relativistic ejecta: jets ....................................................................... 286
      6.7.1 Jet breaks ......................................................................................................... 286
      6.7.2 GRB beaming and energetics .......................................................................... 292
      6.7.3 Orphan afterglows ............................................................................................ 296
      6.7.4 The structure of the jet ..................................................................................... 297
      6.7.5 Polarization: a tool to distinguish the different jet models ............................. 301
   6.8 Inhomogeneous external medium: the case of a wind ............................................ 306
      6.8.1 Inhomogeneities in the circum-burst medium ................................................. 308
   6.9 Variable energy injection ....................................................................................... 311
   6.10 The fireball model confronts observations ............................................................ 314
   6.11 Conclusion ............................................................................................................. 316
   6.12 References ............................................................................................................ 319

7 Early afterglows: the Swift revolution ....................................................................... 331
   7.1 The steep decay and its origin ................................................................................. 333
      7.1.1 The curvature effect ......................................................................................... 333
      7.1.2 The time zero point ($t_0$) effect ....................................................................... 335
      7.1.3 Other effects .................................................................................................... 336
   7.2 The shallow phase .................................................................................................. 337
      7.2.1 Refreshed shock: Energy injection .................................................................. 338
      7.2.2 Different structural jet models ........................................................................ 340
      7.2.3 Prior activity ..................................................................................................... 343
      7.2.4 Varying microphysics parameters .................................................................... 343
      7.2.5 Other models ................................................................................................... 346
      7.2.6 New ideas: the role of the reverse shock ......................................................... 346
   7.3 Revisiting the efficiency of internal shocks ............................................................. 350
   7.4 X-ray flares ............................................................................................................ 353
      7.4.1 Models which have difficulties ....................................................................... 353
7.4.2 Flares and the late central engine activity ..................... 356
7.4.3 The possible origin of the late central engine activity ....... 359
7.5 Chromatic X-ray breaks .................................................. 364
7.6 The jet breaks after Swift ............................................... 367
7.6.1 The jet break ......................................................... 368
7.7 Conclusion ............................................................... 376
7.8 References ............................................................... 377

8 Progenitors ........................................................................... 385
8.1 Introduction ................................................................. 385
8.2 The collapsar model ...................................................... 386
8.3 Collapsars, constraints on the progenitor star .................... 397
8.4 The supranova model ..................................................... 401
8.5 Pulsar models ............................................................... 405
8.6 Observational evidence linking GRBs with the collapse of massive stars ......................................................... 410
8.6.1 GRB–SN association ..................................................... 410
8.6.2 Bumps in optical afterglow light-curves ...................... 415
8.6.3 Additional SN signatures and questions .................... 419
8.7 Neutron star–neutron star mergers ................................... 421
8.7.1 Mergers of two neutron stars ...................................... 422
8.7.2 NS–NS coalescence: the first generation of models ....... 427
8.7.3 NS–NS coalescence: the second generation of models .... 435
8.7.4 NS–NS coalescence: the role of the NS magnetic field and its evolution ......................................................... 442
8.8 Mergers involving a black hole ........................................ 444
8.8.1 Mergers of a black hole and a neutron star .................. 444
8.8.2 Black hole–He star mergers ........................................ 450
8.8.3 Black hole-WD mergers and neutron star-WD mergers . 450
8.9 the galactocentric distribution of compact star mergers ....... 452
8.10 Some recent results on short GRBs and the merger model .... 455
8.11 Conclusion ................................................................. 462
8.12 References ................................................................. 463

9 Perspectives in gamma-ray burst science ................................ 477
GRBs and the early Universe .................................................. 478
9.1 Cosmic rulers ............................................................... 478
9.2 Cosmological probes ..................................................... 485
9.2.1 GRBs and the first stars ............................................. 485
9.2.2 GRBs and quasars ...................................................... 489
9.2.3 Afterglow spectroscopy: absorption lines of the interven-
ing matter ................................................................. 491
9.2.4 Soft X-ray absorption ................................................ 498
9.2.5 Fine-structure transitions ......................................... 501
9.3 Host galaxies ............................................................... 502
Cosmic rays, neutrinos, gravitational waves .............................................. 506
9.4 High-energy photons ................................................................. 506
  9.4.1 Brief summary of high-energy gamma-ray observations ............ 506
  9.4.2 High-energy gamma-ray observatories .................................. 509
  9.4.3 High-energy gamma-ray emission from internal and external shocks ........................................... 511
  9.4.4 High-energy gamma-ray emission involving external shocks ........................................... 512
  9.4.5 High-energy gamma-ray emission involving internal shocks, reverse shocks or X-ray flares ............. 516
  9.4.6 The role of neutrons .......................................................... 517
  9.4.7 The interaction of high-energy gamma-rays in the source ........ 518
  9.4.8 The interaction of high-energy gamma-rays with the intergalactic medium ................................... 519
9.5 Ultra-high-energy cosmic rays ..................................................... 521
9.6 High-energy neutrinos ............................................................. 524
9.7 Gravitational waves ............................................................... 528
Conclusions ...................................................................................... 531
9.8 Some problems encountered by present GRB models ................. 531
  9.8.1 Origin and structure of the magnetic fields ............................ 531
  9.8.2 Particle acceleration and the spectral index of the electron energy distribution ................................... 533
  9.8.3 The microphysics parameters ................................................. 535
  9.8.4 \( \Gamma \), the bulk Lorentz factor of the fireball ......................... 537
  9.8.5 The density of the circum-burst medium ................................. 537
  9.8.6 The early X-ray afterglow: problems posed by the fireball model ........................................... 538
  9.8.7 The progenitors ................................................................. 540
  9.8.8 The standard fireball model .................................................. 543
9.9 Perspectives .................................................................................. 544
  9.9.1 Future instruments .............................................................. 546
9.10 References .................................................................................. 548

Index .................................................................................................... 565
Preface

The Editor proposed us to write this book in May 2004, few months before the launch of Swift. After some thoughts we responded positively to this request because we were really enthusiastic about relating the remarkable GRB story and providing in a single place basic observational and theoretical facts about gamma-ray bursts. The book is organized along these two lines.

Chapters 1 to 4 relate the GRB story in chronological sequence. They describe how observational progress have led a serendipitous discovery to become important domain of astrophysics, with the understanding that gamma-ray bursts are powerful stellar explosions producing a short-lived ultra-relativistic jet directed towards us. Chapters 5 to 8, discuss our basic understanding of gamma-ray bursts. The theoretical description of GRBs involves the evolution of relativistic fireballs and their impact on their environment, and the physical processes at work in astrophysical jets. These questions are addressed in Chapters 5, 6, and 7. Chapter 7 also shows the crucial need for multi-wavelength (X-rays to radio) time-resolved observations, and how the scarcity of these observations currently limits our understanding of gamma-ray bursts. Chapter 8 is about the GRB sources. It discusses exploding stars, like the collapsars leading to newly born fast-rotating black holes (and possibly short-lived magnetars), and various types of compact star mergers. It mentions some methods to distinguish between these two types of progenitors, and their difficulties. Chapter 9 briefly explore the possible future developments of the field. It explains the connection of GRBs with many present-day astronomical issues, like the history of star formation, the production of cosmic rays, and the detection of gravitational waves, and comments on how these events are participating in the better understanding of our Universe. ‘Gamma-Ray Bursts’ is intended for scientists and students who are interested in the main concepts of the quickly evolving field of GRB studies. It describes the main observational and theoretical work which have led to our present understanding, and it provides considerable bibliography. The curious reader (not necessarily scientist) may also be interested in the GRB story starting with the
serendipitous discovery of these events in Chapter 1 and ending in Chapter 9 with a discussion of the many connections of GRBs with other fields of astrophysics.

When we started this venture, we were aware of three major difficulties of the task: our English was poor, the field was evolving very quickly (at least as fast as we could write!), and it was not possible in a single book to address all the astrophysical issues connected with GRBs. Regarding the first point we ask for the indulgence of the reader. The other points explain why we have focused this book on GRBs observations, theory and progenitors, with few excursions outside these issues (in Chapter 9). We have also tried to stay on firm grounds (as far as we can judge), avoiding going into the details of questions which remain speculative, but trying to list them. Moreover, we rapidly realized another complication which we had not anticipated: we had to face a vast bibliography preventing us from completeness, even with a total of more than 1700 references listed at the end of each chapter. We have attempted to provide a set of references sufficient to allow the detailed exploration of the subjects tackled in the book, but we definitely apologize for the references that have escaped our attention. We have used some reviews extensively, when this is the case this is mentioned in the text, and we take advantage of this preface to thank the authors of these reviews.

We have included in the book various features that could help the reader. The chapters are mostly independent and their content is outlined in the introduction. The last section of each chapter is a short summary of its content. Cross-references between sections when this is required. An index of the main concepts is given at the end of the book, and, following a long-lasting tradition among GRB researchers, this index also contains references to the individual GRBs mentioned in the book, allowing anyone to search where and why we have mentioned his/her favourite GRB.

Finally, many thanks are due to Kevin Hurley who has accepted to go through the manuscript and has corrected many mistakes. We fully endorse the remaining errors and inaccuracies and we apologize for them. We also thank the Editor for offering us the opportunity to present the richness of a scientific field that we love and to tell the story of a serendipitous discovery that became in a quarter of a century an extremely active domain of research, contributing to a better understanding of our Universe.
Acknowledgments

*Gamma-Ray Bursts* is the result of a collective effort and we would like to thank here all those who have contributed to bring this book to life.

First, Clive Horwood, Publisher, Praxis, whose confidence put us on the way in May 2004, while we were far from realizing the size of the task (“The journey of a thousand miles starts with a single step”—Lao Tseu). Clive was also very patient and comprehensive when we accumulated delays in chapter delivery. One benefit of these delays has been the possibility to fully take into account the beautiful discoveries of Swift. By chance, and in contrast with the predictions of some of our colleagues, they were not long enough to include the first results of the nicely working GLAST/Fermi mission.

We are also grateful to the referees who recommended this project, with excellent advice on the opportunity and content of the book, and to the colleagues who have read and corrected parts of the book, especially A. Blanchard. During the production process, we have discovered a number of persons who all contributed to the realization of the book. Dr John Mason, the Science Editor, Mike Shardlow, the Copy Editor, and Neil Shuttlewood, the Typesetter, who took into consideration all our requests during the stage of proof corrections, we warmly acknowledge their work. Many thanks are also due to our families who supported us during this long, and sometimes difficult, project. Finally we are indebted to the GRB community at large whose work is at the centre of this book. *Gamma-Ray Bursts* is a tribute to all the researchers and engineers who have contributed to the blossoming of this area of Astrophysics.

Special thanks are due to two colleagues R. Mochkovitch and K. Hurley. R. Mochkovitch, a well-known expert in GRB theory, kindly supervised the theoretical chapters of the book on GRB models and progenitors. K. Hurley, who started his research on γ-ray bursts at the CESR in Toulouse 30 years ago, is an active GRB scientist and a very regular and reliable partner. He has accepted to revise the nine chapters, which is a considerable amount of work, and an invaluable contribution to the book. These two colleagues contributed significantly to improve the quality of the book.
Foreword

The history of cosmic gamma-ray bursts, although relatively short by astronomical standards, has been a very eventful one. In the space of a few decades, they have gone from a curiosity, thought to be related to small energy releases involving Galactic neutron stars, to the most powerful cosmological explosions known. At the same time, they have become tools for the study of stellar evolution, supernovae, galactic structure, the intergalactic medium, and relativistic plasmas, to name only a few subjects. Well over 9,000 papers have now been published on gamma-ray bursts, and a new one appears every day or so, a rate which is about equal to that of the detection of new bursts. The authors of this book have both witnessed and been directly involved first-hand in this remarkable story. Gilbert Vedrenne was there at the beginning; as vice-director of the CESR (Toulouse), he was in a unique position to put the resources of the laboratory to work in studying gamma-ray bursts by building some of the first dedicated satellite experiments, which were launched on Russian spacecraft. Jean-Luc Atteia arrived on the scene only slightly later as a graduate student, and began his career by analyzing the resulting data. Both authors have continued to be deeply involved in the unfolding story over the years.

Although other books have now been written on the subject of gamma-ray bursts, this one is unique. It starts by tracing the history of the subject in detail, and then delves into the complex physics of the phenomenon. The first part of this book can be read as a history of the subject by anyone with a minimum of scientific training. The second part is suitable for advanced undergraduates and graduate students, and will also be useful as a reference for researchers in the field. It arrives at a moment when the major mystery—burst distances—has been resolved, and a flood of new data is demanding ever-more clever and detailed explanations, and forcing us to abandon some old preconceptions. Gamma-Ray Bursts captures the excitement, the changing ideas, and the cutting-edge physics of an extraordinary astrophysical phenomenon.

Kevin Hurley
November 2008
J-L.A dedicates the book to Sylvie
and
G.V. dedicates the book to his family
Figures

(figures marked with an asterisk also appear in the color section)

1.1 The GRB discovery: the first light curves of GRB 700822 recorded at three Vela satellites .......................................................... 2
1.2 Light-curves of GRBs: some examples of their diversity ................. 3
1.3 The possible bimodal distribution of GRB durations was evidenced by the first experiments ......................................................... 4
1.4 Narrowing of the width of the peaks with energy in GRB 830801 ........ 5
1.5 Spectra of nine GRBs observed with Imp-7 .................................... 7
1.6 An example of a GRB spectrum (GRB 720427) obtained with the X-ray and gamma-ray spectrometers on Apollo 16 ..................... 8
1.7 Spectral evolution in a well-structured GRB 781119 observed by SIGNE experiments on Venera 12 and 11 ....................................... 10
1.8 Spectral evolution of pulse structures in GRB 821104 observed with Solar Maximum Mission satellite ................................. 11
1.9 A typical GRB spectrum illustrating the non-thermal nature of the emission and the peak power evolution during the burst ................ 12
1.10 Some examples of possible cyclotron lines seen by KONUS experiments on Venera 11 and 12 probes ............................................ 13
1.11 Possible cyclotron features in the spectra of three GRBs observed with Ginga ................................................................. 15
1.12 A generic GRB spectrum with its continuum and possible features .... 16
1.13 The triangulation method which has been largely used to localize GRBs ... 17
1.14 Maps of GRBs localized by KONUS and IPN experiments ............ 19
1.15 This is the schematic illustration realized by Briggs (1995) of the model constraints imposed by observing the 2D angular distribution and the intensity distribution of GRBs ........................................... 22
1.16 The log N–log S distribution which is the number of bursts per year N(> S) with fluence greater than S (erg cm^{-2}) as a function of S ................... 23
1.17 Burst fluence as a function of the burst duration indicating the selection effects for long bursts with low fluences ........................... 25
1.18 Cumulative size frequency distribution of the KONUS bursts versus fluence, peak energy flux, and peak counting rate .......................... 26
1.19 Two examples of well-localized GRBs, with the IPN GRB 781119, and GRB 790613 ........................................ 32
1.20 An artist’s view of the thermonuclear model of GRBs, at a time when GRBs were supposed to be of Galactic origin ........................................ 38
1.21 Schematic diagram showing the physical regions likely to be responsible for the continuum and line features of GRB spectra ........................................ 39
2.1 An example of a very short burst seen by BATSE .......................... 50
2.2 Distribution of $T_{90}$ for 427 GRBs from the 3rd BATSE Catalog .................. 52
2.3 Distribution of hardness ratio versus $T_{90}$ for 222 GRBs, and the same distribution obtained with Phebus detectors on the Granat mission ............. 53
2.4 An example of a GRB spectrum extending to $\sim$10 MeV, and GRB 940217 observed by Egret, Ulysses and BATSE ............................... 54
2.5 Sky distribution of the 1005 BATSE GRBs in an Aitoff–Hammer projection in galactic coordinates regardless of the trigger energy range ......................... 58
2.6 Integral log $N$–log $P$ distribution for the standard 50–300 keV trigger energy range for peak flux measured on a timescale of 1024 ms .............................. 60
3.1 Images of the source 1 SAX J0501.7+1146, detected with the BeppoSAX Medium Energy Concentrator Spectrometer ........................................ 90
3.2 Variation with time of the X-ray flux from GRB 970228, in the 2–10 keV energy range ........................................ 91
3.3 V-band images of a 1.5′ × 1.5′ region of the sky which contains the optical transient (OT) associated with GRB 970228 ........................................ 92
3.4 Light-curves of the afterglow of GRB 970228 at various wavelengths, from X-rays to near-infrared, and the afterglow of GRB 970228 measured in the V, Rc, and Ic bands at late times ........................................ 93
3.5 R-band light-curve of the optical transient associated with GRB 970508 . . . . . 94
3.6 X-ray light-curve (2–10 keV) of GRB 970508 and its afterglow (WFC and NFI) 95
3.7 The X-ray to radio energy spectrum of GRB 970508, measured 12.1 days after the burst ........................................ 97
3.8 Schematic diagram showing the size of the visible patch of a relativistic shell 103
3.9 Light-curves of the optical afterglow of GRB 990510 measured at different wavelengths ........................................ 105
3.10 Observed distribution of jet break times and jet opening angles .................. 106
3.11 Distribution of the apparent isotropic energy $E_{iso}$, for GRBs with known redshifts, and the distribution of the beaming corrected energies .................. 107
3.12 R-band light-curve of the afterglow of GRB 980326, and another example of a possible SN contribution in the light-curve of GRB 011121 .................. 111
3.13 The X-ray spectrum of the afterglow of GRB 011211 at redshift $z = 2.14$, measured with XMM-Newton ........................................ 115
3.14 The light-curve of GRB 920723 observed with Sigma in the 35–500 keV energy range ........................................ 118
3.15a The optical light-curve of GRB 990123 ........................................ 120
3.15b The R-band light-curve of the afterglow of GRB 990123 extended to 10 days 121
3.16 Correlation between $E_{peak}$ and $E_{iso}$, the isotropic total radiated energy .......................... 122
4.1 The HETE-2 satellite during final tests ........................................ 136
4.2 The network of secondary ground stations receiving the HETE-2 VHF signal 137
4.3 HETE light-curves of GRB 030329 ........................................ 142
4.4 Optical afterglow of GRB 030329 in B, V, R, I bands, and the X-ray-to-radio energy spectrum of GRB 030329, measured 0.5 day after the burst ....... 143
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>Evolution of the GRB 030329–SN 2003dh optical spectrum, and the MMT spectrum measured on April 8 with continuum subtracted</td>
</tr>
<tr>
<td>4.6</td>
<td>Gamma-ray light curve of GRB 050709 showing the short GRB followed by a long, soft bump, and Hubble Space Telescope images showing the fading afterglow and host galaxy of GRB 050709.</td>
</tr>
<tr>
<td>4.7</td>
<td>X-ray images of GRB 031203 observed with Epic on XMM-Newton and covering the 0.7–2.5 keV energy range</td>
</tr>
<tr>
<td>4.8</td>
<td>Schematic view of the Swift satellite</td>
</tr>
<tr>
<td>4.9</td>
<td>The main elements of the Swift platform</td>
</tr>
<tr>
<td>4.10</td>
<td>Cumulative distribution of GRBs as a function of redshift for 43 pre-Swift bursts and for 19 Swift bursts.</td>
</tr>
<tr>
<td>4.11</td>
<td>Examples of BAT and XRT light-curves for some Swift GRBs</td>
</tr>
<tr>
<td>4.12</td>
<td>The beautiful light-curve of the X-ray afterglow of GRB 050315</td>
</tr>
<tr>
<td>4.13</td>
<td>A cartoon X-ray light-curve based on the observational data from the Swift XRT, and another schematic view of the early X-ray light-curve with the two types of afterglows</td>
</tr>
<tr>
<td>4.14</td>
<td>Light-curve of the X-ray afterglow of GRB 050502B in the energy range 0.2–10 keV, and fit of the X-ray flare and the underlying decay</td>
</tr>
<tr>
<td>4.15</td>
<td>Several examples of X-ray flares</td>
</tr>
<tr>
<td>4.16</td>
<td>Scatter plot of X-ray flare parameters</td>
</tr>
<tr>
<td>4.17</td>
<td>Spectrum of GRB 060218/SN 2006aj taken with the 6.5-m Multiple Mirror Telescope</td>
</tr>
<tr>
<td>4.18</td>
<td>XRT and UVOT light-curves of GRB 060218</td>
</tr>
<tr>
<td>4.19</td>
<td>Evolution of the temperature and radius of the soft thermal component of GRB 060218</td>
</tr>
<tr>
<td>4.20</td>
<td>The early X-ray afterglow of GRB 050904</td>
</tr>
<tr>
<td>4.21</td>
<td>Optical and X-ray light-curves of GRB 050904 showing the end of the prompt emission and the transition to the afterglow</td>
</tr>
<tr>
<td>4.22</td>
<td>Light-curves of the afterglow of GRB 061007 in all six UVOT filters</td>
</tr>
<tr>
<td>4.23</td>
<td>The light-curve of GRB 060614 observed with BAT in four energy channels</td>
</tr>
<tr>
<td>4.24</td>
<td>Light-curves of supernovae SN 1998bw, SN 2002ap, and SN 2006aj as they would have appeared at the redshift of GRB 060505 and at the redshift of GRB 060614</td>
</tr>
<tr>
<td>5.1</td>
<td>Schematic behavior of the jet Lorentz factor $\Gamma$ in the different regions that can be identified in the standard model</td>
</tr>
<tr>
<td>5.2</td>
<td>The different timescales from a relativistic expanding shell with a width $\Delta$ and a Lorentz factor $\Gamma$</td>
</tr>
<tr>
<td>5.3</td>
<td>This figure illustrates the meaning of $t_{\text{shell}}$ for spherical relativistic shells and how it can be very simply obtained</td>
</tr>
<tr>
<td>5.4</td>
<td>Synchrotron emission from electrons accelerated in a relativistic shock</td>
</tr>
<tr>
<td>5.5</td>
<td>Simulated burst profiles for three initial distributions of the Lorentz factor in the ejecta</td>
</tr>
<tr>
<td>5.6</td>
<td>Low-energy power-law spectral index of time-resolved spectra of bright bursts detected with BATSE and BeppoSAX</td>
</tr>
<tr>
<td>5.7*</td>
<td>Schematic illustration of the electromagnetic model</td>
</tr>
<tr>
<td>6.1*</td>
<td>An artist’s view illustrating the different steps of the basic standard model with the internal and external shocks and the different radiations they emit</td>
</tr>
<tr>
<td>6.2*</td>
<td>Schematic view of the various phases involved in the fireball model with internal and external shocks</td>
</tr>
</tbody>
</table>
6.3 Schematic representation of the regions that can be identified when a relativistic fireball interacts with the interstellar medium, and the Lorentz factor (\(\gamma\)), density (\(\rho\)), and pressure (\(\rho v^2\)) in the four regions derived from simulations .

6.4 Synchrotron spectra produced by a relativistic shock with a fast-cooling power-law electron distribution .

6.5 Synchrotron light curves ignoring self-absorption in the two cases: high frequency and low frequency .

6.6 The optical light-curve of GRB 021211 starting \(\sim 90\) seconds after the burst

6.7 Optical light-curve of GRB 021004 .

6.8 Comparison of prompt gamma-ray and prompt optical-light curves of GRB 041219 and the famous GRB 990123.

6.9 Schematic optical light-curve for a synchrotron dominated fireball in a dense environment, such as a wind or a dense interstellar medium .

6.10 Schematic illustration of the equal arrival time surface, the surface from where the photons emitted at the shock front arrive at the same time to the observer.

6.11 Shape of the jet break characterized by the evolution of the temporal decay index as a function of the observed time.

6.12 Diagram of the emission from a uniform relativistic jet with sharp edges and half opening angle \(\theta_0\) .

6.13 Beaming-corrected GRB energies .

6.14 Distribution of various energy measurements for 18 GRBs .

6.15 Constraints on the rate of visible afterglows from GRBs obtained with the CFHTLS Very Wide Survey .

6.16 Radio to X-ray light-curves of the afterglow of GRB 030328 .

6.17 Polarization of the optical afterglow of GRB 020813.

6.18 Optical polarization measurements of five GRB afterglows up to 2002 .

6.19 Optical light-curve resulting from the presence of a Gaussian over-dense region in the ISM .

6.20 Broadband signature (X-ray, optical, radio) of the injection of a kinetic-energy-dominated shell .

7.1 The phases of the afterglow of GRB 050315.

7.2 Effect of \(l_0\) on a GRB light-curve .

7.3 Fit of the X-ray light-curve of GRB 050315 with a Gaussian jet and a ring-shaped jet uniform within \([\theta_c; \theta_c + \Delta\theta]\) .

7.4 X-ray light-curves from a Gaussian jet seen off-axis .

7.5 Angular distribution of the kinetic energy per solid angle of the ejecta in a ‘double outflow, and tentative fit of the X-ray light-curve of GRB 050315 with a two-component model’ .

7.6 Synthetic X-ray light-curves in the 0.3–10 keV range for different reverse shock models .

7.7 Stratification function \(\Gamma_{ej} (E_{ej})\) in the model of Uhm and Beloborodov (2007) .

7.8 Blastwave evolution in the reverse shock model of Uhm and Beloborodov (2007) .

7.9 The three-stage GRB engine proposed by Staff, Ouyed, and Bagchi (2007) .

7.10 Illustration of the mechanism leading to the flattening of the X-ray afterglow light-curve and to X-ray flares in the three-stage model of Staff, Ouyed, and Bagchi (2007) .

7.11 General structural features of the inner MHD flow at three different accretion stages.
7.12 Contours of the mass–density 22.3 ms after the merger of two neutron stars, and a sample of fall-back trajectories resulting from the double NS merger 363
7.13 Light-curves of six Swift GRB afterglows showing a chromatic X-ray break, not seen in the optical 365
7.14 Examples of X-ray light-curves of Swift GRB afterglows fit with a prompt emission component and an afterglow component 370
7.15 X-ray and optical (UVOT) light-curves of GRB 050525A 371
7.16 Another example of a possible achromatic break at $T \sim 10^5$ seconds in the afterglow of GRB 060526 372
7.17 Examples of potential jetbreaks at $\sim$1 day 373
7.18 The beautiful X-ray light-curve of XRF 050416A 374
7.19 X-ray light-curves in which a jet break is not seen at the epoch predicted by the Ghirlanda relation 375
8.1 Three scenarios of collapsar formation proposed by Fryer, Woosley and Hartmann (1999), involving either a single star or binary stars 388
8.2 Binding energy and explosion energy of a stellar core as a function of the mass of the progenitor 390
8.3* Lorentz factor for the collapsar model $W_2$ of Zhang, Woosley, and MacFadyen (2003) 394
8.4 Schematic diagram illustrating the propagation of a relativistic jet through the stellar envelope of the progenitor 396
8.5 Evolution of the optical luminosities and absolute magnitudes of seven Type Ic supernovae 414
8.6 Late R band light-curve of the afterglow of GRB 041006 417
8.7 Rest-frame peak optical magnitude of GRB-associated SNe and local SNe Ibc as a function of the mass of $^{56}$Ni synthesized in the explosion; and histogram of peak optical magnitudes for GRB/XRF associated with SNe and local SNe Ibc 418
8.8 Two scenarios considered by Fryer, Woosley, and Hartmann (1999) for double (DNS) mergers with the final production of a GRB 423
8.9 Two scenarios considered by Fryer, Woosley, and Hartmann (1999) for BH–NS mergers with the final production of a GRB 424
8.10 Two scenarios considered by Fryer, Woosley, and Hartmann (1999) for BH–WD mergers with the final production of a GRB 425
8.11 Illustration of the mechanisms which could produce a GRB from a NS merger 428
8.12 Log(density–$g$ cm$^{-3}$) along the rotation axis ($z$-axis) of the merged remnant of two 1.4 $M_\odot$ neutron stars in corotation. Log(energy deposition—erg s$^{-1}$ cm$^{-3}$) via $\nu\nu$ annihilation for the same simulation 429
8.13* Ratio of the energy deposited via neutrino annihilation to baryon rest mass energy for the region above the poles of the merged remnant 430
8.14 Magnetic field development in a DROCO 434
8.15 Neutrino luminosity $L_\nu$ for disk masses 0.3 and 0.06 $M_\odot$ and for different viscosities, and luminosity $L$ when the outflow is powered by the Blandford–Znajek mechanism 437
8.16 Summary proposed by Shibata and Taniguchi (2006) giving the outcome after the merger of two neutron stars 440
8.17* Snapshots of the coalescence of two magnetized NS followed between 1.83 ms and 11.34 ms 443
8.18 Initial, early, and intermediate configuration of a NS merging with a black hole, and matter configuration at the end of the simulation 449
8.19 Cumulative distribution of the predicted distance of double NS mergers to the center of a massive galaxy, and to the center of a dwarf galaxy............. 454
8.20 Light-curve of the X-ray afterglow of GRB 050724, combining data from BAT, XRT, and Chandra ....................................................... 457
8.21 Histogram of observed R band magnitudes of host galaxies of short GRBs . 460
9.1 Rest-frame peak energy versus isotropic energy and versus collimation corrected energy ......................................................... 482
9.2 Predicted GRB rate (GRBs per year) for Swift as a function of redshift..... 487
9.3 Fraction of GRB–DLA and QSO–DLAs per HI, FeII, and Zn II column–density interval ......................................................... 491
9.4 Depletion pattern in five GRB–DLAs including four or more heavy elements 492
9.5 Distribution of the host galaxy visual extinction in the source frame for the Golden Sample of 19 GRB aferglows .................................. 493
9.6 Dust-to-gas ratios in the host galaxies along the line of sight to the GRBs of the Golden Sample .................................................... 494
9.7 Optical spectrum of the afterglow of GRB 20813 at z = 1.255, after correction for galactic extinction .................................................. 495
9.8 Redshift evolution of metallicity for nine GRB–DLAs and 197 QSO–DLAs. 497
9.9 Neutral hydrogen column density as a function of the X-ray equivalent hydrogen column density ....................................... 500
9.10 Estimate of the total stellar mass content of 32 GRB host galaxies, based on SED fitting to the multi-band optical/NIR photometry ......... 504
9.11 Mass–metallicity relation of seven GRBs and normal star-forming galaxies for redshifts between 0.4 and 1 ............................................ 504
9.12 Evolution of the spectrum of GRB 941017 as a function of time .......... 508
9.13 Point-source sensitivity of various very high-energy instruments ........ 510
9.14 Regions in the ε_e, ε_B parameter space in which various radiation mechanisms dominate at selected frequencies ............................ 513
9.15 Gigaelectronvolt fluence versus time for three types of bursts starting from τ_{dec} and ending at the time when the bulk Lorentz factor Γ = 2 .............. 514
9.16 Optical depth for γ-rays above 50 GeV for redshifts between 0.03 and 0.3 . 520
9.17 Diffuse muon–neutrino flux shown as solid lines from sub-stellar jet shocks in two GRB progenitor models ........................................... 525
9.18 The ratio ε_e/ε_B for 10 GRBs analyzed by Panaitescu 2006 .................. 536
9.19 Photograph of the participants at the symposium ‘Cosmic Gamma-Ray Bursts’ held in Toulouse, France, in November 1979 ......................... 547
Tables

4.1 Main characteristics of BAT on Swift ........................................... 162
4.2 Main characteristics of XRT on Swift ............................................ 164
4.3 Main characteristics of UVOT on Swift ......................................... 165
6.1 $\alpha$ and $\beta$ for fast cooling into a constant density ISM ............. 274
6.2 $\alpha$ and $\beta$ for slow cooling into a constant density ISM ................ 274
6.3 Temporal index $\alpha$ and spectral index $\beta$ in various afterflow models .... 291
6.4 The results of the 25 polarization measurements performed on 9 GRBs before January 2003 ................................................................. 305
8.1 GRB–SN associations up to mid-2005 ............................................. 411
8.2 Approximate results for various evolutionary scenarios. ................. 451
Introduction

The fascinating gamma-ray burst (GRB) story started in 1973 with the report of the discovery of unexpected bursts of gamma-rays detected by military satellites. It then took decades before this topic of the young high-energy astronomy field became an important area of astrophysics, involving a large and very active scientific community. Having participated in the birth of this field in the 1970s and 1980s, we were enthusiastic when the editor proposed that we should relate the fabulous story of these gamma-ray bursts. We say fabulous because it presents all the ingredients of a fortuitous discovery which, after a long period of gestation, became a major scientific topic. Naturally, we accepted this demand even though the work was considerable because of the multiple developments of this domain, brought about by the continuous succession of GRB missions, particularly the very successful Swift satellite, still in orbit. Without being exhaustive we have tried to present the major milestones of the GRB saga and our understanding of this field in 2008.

The GRB saga started with an exciting pioneering period marked by an explosion in the number of bursts observed with a growing number of satellites; but there was no decisive progress in discovering the real sources of these explosions. In the 1980s, nearby galactic neutron stars were considered to be the most probable candidates to explain GRBs, by analogy with the models explaining most of the galactic X-ray sources. In these early times, the number of models was comparable with the number of bursts. We describe in Chapter 1 the long period of investigation that was dedicated to the study of the temporal and spectral characteristics of gamma-ray bursts. As they were unpredictable in time and direction, the best way of detecting a large number of them was to use omnidirectional gamma-ray detectors. The determination of GRB characteristics included the measurement of their light-curves—to search for possible periodicities—and the study of their spectra—to establish if they were thermal or non-thermal and to search for possible spectral features. After a few years a large sample of bursts was available, often obtained easily, with very different gamma-ray detectors working in space above ten or a few tens of kiloelectronvolts. But these data
were insufficient to fully understand the physics of GRBs and to identify the sources of these emissions. Some similarity could be found with X-ray bursts seen at lower energy and already associated with accreting neutron stars in our galaxy. But this analogy did not answer the question of the origin of GRBs.

The need to localize and identify the sources of these sporadic emissions was very much present from the beginning, but this was difficult as the sources did not repeat and occurred randomly on the sky. The accurate timing of GRBs detected with omnidirectional detectors on at least three satellites had appeared very early to be the best way to localize GRB sources with the best precision: this is the triangulation method. Another method was also used to provide coarse GRB positions: the detection of GRBs with several large-field-of-view detectors on the same satellite with different orientations covering the entire sky. Very quickly, using this last technique, an all-sky map of the sources was obtained by the Leningrad group led by E. Mazets. The GRB distribution appeared to be isotropic, without the concentration in the Galactic Plane that was observed for most Galactic populations. This result led to the conclusion that the sources were Galactic but in our close proximity, with a distance scale smaller than the scale height of the Galactic Disk. The triangulation method used for another sample of GRBs led to similar results. This isotropy was, however, puzzling for Galactic sources and M. Schmidt proposed using the $\langle V/V_{\text{max}} \rangle$ test to measure the homogeneity of the sources in the radial direction. This test computes, for each GRB, its relative position within the volume accessible to the instrument. For homogeneous sources one expects $\langle V/V_{\text{max}} \rangle = 0.5$. The results for GRBs were marginally below 0.5, suggesting a deficit of faint bursts. At that time the apparent contradiction between the isotropy and the non-homogeneity of the sources was not considered to be too dramatic, because localizations were usually obtained for the brightest GRBs seen by several detectors, while the $\langle V/V_{\text{max}} \rangle$ test involved all GRBs, including the faintest. Most researchers were expecting that more sensitive instruments would find faint sources concentrated in the Galactic Plane. Some of them, however, considered this isotropy as a sufficiently serious constraint to propose extragalactic and cosmological models. This possibility was defended, for instance, by B. Paczyński and his colleagues from the middle of the 1980s. In the end, the searches for counterparts at all possible wavelengths—optical, radio, and X-rays—using the best localizations obtained by triangulation, were all negative, conserving the mystery of the nature of these sources.

Hence after more than 15 years (1973–1991) of investigation a large sample of GRBs was accumulated with a large diversity of instruments on many satellites, often dedicated to X-ray or gamma-ray astronomy, and on planetary probes. They made it possible to collect a large sample of light-curves with rapid variability at the level of milliseconds, but with no convincing periodicities. When measured, the time-resolved spectra during the burst were often quite variable. Moreover lines were detected at $\sim$400 keV, interpreted as originating from the annihilation of $e^+/e^-$ at 511 keV, redshifted to 400 keV by the gravitational field of the compact source. Absorption lines at 20–40 keV were also observed by the KONUS instrument and convincingly confirmed by Ginga for a few GRBs. These lines were taken as an indication that GRBs were produced by strongly magnetized neutron stars (NS). These features were
indisputably considered to be strong arguments for the association of GRBs with nearby Galactic neutron stars, isolated or in binary systems. The absence of counterparts for sources situated at few hundred parsecs (to accommodate the observed isotropy) was compatible with isolated NS or with NS in a binary system with a faint companion and a low accretion rate. The energy involved in GRBs was reasonable ($10^{38}$ erg) due to the proximity of the sources and the sources had to emit $10^3 - 10^5$ bursts to explain the observed rate of a few $\times$ 10$^2$ GRBs per year. So by the end of this period NS models had been developed which explained the sources of GRBs reasonably well. The difference between GRBs and X-ray bursts was attributed to the lower accretion rate of GRB sources possibly accompanied by a higher magnetic field. It is finally interesting to note that the issues that would become crucial for cosmological models were already being addressed at that early time: photon–photon opacity and the non-thermal shape of the spectrum, fast variability, the nature of the radiation mechanism, and so on.

In 1991 NASA launched an ambitious mission, called the Compton Gamma-Ray Observatory (CGRO), which was quite successful. CGRO was so big (17 tons) that it had to be launched from the space shuttle. Planned for launch in 1986, it was delayed until 1991 after the tragic Challenger accident in January 1986. CGRO carried four science instruments, among them BATSE which improved the sensitivity to GRBs by about an order of magnitude. We analyze the major breakthroughs made possible by BATSE on CGRO in Chapter 2. Thanks to the sensitivity and to the long operating time of BATSE an unprecedented number of bursts (nearly 3000) was observed with the same instrument. The major properties of GRBs: duration, rapid variability, lack of periodicities, existence of two classes of GRBs, short and long, were completely confirmed. The non-thermal nature of their energy spectra and the spectral variability within a burst were confirmed and well studied. On the other hand, one very important result which had strongly influenced the neutron star paradigm described in Chapter 1, the presence of absorption lines at 20–40 keV, was not confirmed in any of the GRB spectra.

The isotropy of the angular distribution already observed before BATSE was quickly confirmed for hundreds of crudely localized GRBs. The sensitivity of the BATSE detectors, which allowed sampling GRBs five to six times more distant than previous detectors (a volume 100 to 200 times larger) and the large size of the sample made it possible to demonstrate the isotropy of weak GRBs. The $\langle V/V_{\text{max}} \rangle$ test performed on BATSE GRBs demonstrated without any doubt a deficit of faint sources. The conjunction of the indisputable isotropy of the GRB distribution and the non-homogeneous distribution of these bursts is the most remarkable result of BATSE. This result presented decisive arguments for rejecting a GRB origin linked to nearby neutron stars in the Galactic Disk. This had dramatic consequences for the GRB energy scale. For a disk population at typically 100 pc the energy requirement per burst was $10^{38}$ erg. With BATSE this possibility was rejected and the minimum distance scale compatible with the isotropy of the sources was 100 kpc, corresponding to an extended corona around our Galaxy. In this case the energy released reaches $\sim 10^{44}$ erg. But even extended Galactic Halo models were having difficulties with BATSE results. Hence the idea that GRBs could be at cosmological distance (a
few gigaparsecs) with enormous energies liberated in the GRB sources (beyond $10^{52}$ erg) was increasingly accepted.

Cosmological sources would obviously explain the isotropy of the angular distribution, the non-homogeneous distribution of sources and the lack of quiescent counterparts. But the non-thermal nature of the emission and the energy released remained crucial issues. The problem of cosmological models was the very large $\gamma\gamma$ opacity in the source, which would quench high-energy emission. Fortunately solutions were proposed early on involving the relativistic expansion of the source (the fireball). In the absence of counterparts many authors studied the statistical properties of the large GRB sample provided by BATSE in detail. The main purpose of these studies was the search for the expected consequences of the expansion of the Universe on cosmological GRBs: time dilation, spectral softening, and the flattening of the intensity distribution. These studies compared bright and faint GRBs, assuming that the faintest GRBs were on average the most distant. Unfortunately the intrinsic luminosity function of GRBs is so broad (and the range of redshift comparatively small, about a factor of 10) that it was not possible to disentangle cosmological effects from intrinsic properties convincingly.

Finally, despite the beautiful results of BATSE, the absence of counterparts at all wavelengths did not allow GRB sources to be identified. The major questions asked by B. Paczynski and others in 1992 were: How far away are GRBs? What are they? How do they generate gamma-rays?, and they remained without response.

Fortunately a new satellite devoted to X-ray and gamma-ray astronomy was launched in 1996. This Italian–Dutch mission, BeppoSAX, made it possible to detect and locate GRBs, to perform deep X-ray observations of them in the hours following detection, and eventually to discover their X-ray afterglows. This was a major breakthrough because, for the first time, the GRB localizations available in few hours were precise enough to use the largest ground-based telescopes, allowing the GRB afterglow at optical and radio wavelengths to be detected. Hence BeppoSAX solved the GRB mystery: the counterparts of GRBs were found, their distances measured, and their host galaxies identified. These explosions are really at cosmological distances, often at redshift $z = 1$ or more. BeppoSAX opened a brilliant new era for our understanding of GRBs. For these remarkable achievements the BeppoSAX team was awarded the Rossi prize of the American Astronomical Society in 1998 and the Descartes prize of the European Community in 2002.

In Chapter 3 we report the major results of BeppoSAX, which have been so important in elucidating some of the GRB mysteries. First, BeppoSAX made it possible to discover the X-ray afterglow of GRBs, which had been foreseen by theoreticians, and to provide precise GRB localizations in the hours following the burst. The afterglows, which remain detectable for hours and sometimes days, allowed us for the first time to study GRBs with the classical tools of astronomy: large optical and radio telescopes on the ground and X-ray telescopes in space. During the famous ‘GRB year’ 1997 several GRBs were discovered and rapidly followed up by multi-wavelength observations, and their redshifts were measured thanks to optical spectroscopy with large ground-based telescopes, showing that GRBs take place at cosmological distances (i.e. gigaparsecs), with the consequence of a considerable
energy release. At these distances GRBs are the most powerful explosions in the Universe, and over tens of seconds they liberate an energy equivalent to that produced by the Sun during its entire lifetime (\(\sim 10^{10}\) years). In the same year one burst was detected at a large redshift, \(z = 3.42\). The host galaxy was an actively star-forming galaxy which had produced a GRB at a time when the Universe was only 2 billion years old. Later, with Swift, GRBs would be observed at even larger redshifts. We have emphasized the energy problem; it is real because in some cases the GRB energy is found to be comparable with the energy associated with the rest mass of a NS if one assumes isotropic emission. This led very early to the introduction of the possibility of collimation of the GRB emission. If it is significant this can reduce the amount of emitted energy by two or three orders of magnitude, increasing the number of GRBs that have to be produced by the same factor. GRB collimation is extremely important and the ways to measure the collimation angle were widely discussed in this epoch. Achromatic breaks in the afterglow light-curves were expected if the bursts were collimated. These breaks have been observed in many visible afterglows, but the X-ray observations were often too scarce to confirm their achromaticity over a broad range of wavelengths. This critical issue would be reconsidered after the launch of Swift. The breaks observed in the visible afterglows permitted a measurement of the jet opening angle for several GRBs, and measurement of the true energy released in gamma-rays after correction for the beaming factor. Typical beaming angles range from \(1^\circ\) to \(10^\circ\), implying that on average only one GRB out of a hundred or more has its jet directed towards the Earth.

BeppoSAX also provided crucial information on the progenitors of GRBs. The accurate astrometry of GRB afterglows within their host galaxies with the Hubble Space Telescope showed that GRBs occur in star-forming regions of star-forming galaxies. Since these regions harbor young massive stars which probably end up as black holes, models invoking the emission of relativistic jets by newly born stellar black holes were quickly favored. In 1998 BeppoSAX provided the first clue of a possible connection between a supernova (SN) and a GRB. It was noticed that the error box of the long GRB 980425 contained a new supernova SN 1998bw, whose time of explosion, measured to 1–2 day accuracy, was compatible with the time of the GRB. This important observation shed some light on the possible progenitors of GRBs. This SN–GRB connection, which would be confirmed later by HETE-2 and Swift, has been very useful in supporting the collapsar model, which is analyzed in Chapter 8; this model is based on the explosion of a massive star of Wolf–Rayet (WR) type leaving behind a black hole (BH) with a transient accretion disk as the source of the relativistic fireball. This GRB–SN association was an extraordinary revelation. Supernovae were mentioned when the enormous energy associated with a GRB was established, but after the discovery of this possible association GRBs really appeared to be connected with the explosions of massive stars. But of course special conditions were needed because the vast majority of supernovae do not emit the relativistic jet needed to produce a GRB (or at least the jet cannot escape from the star). This is illustrated by the fact that supernovae occur at a rate of about one per second in the observable Universe while GRBs occur at a rate of one per day (or one every 1000 seconds, correcting for beaming).
xxx  Introduction

To fully appreciate the diversity and richness of the BeppoSAX results, the reader should also be aware that this mission has led to the discovery of many new features of gamma-ray bursts:

- The confirmation of the existence of a new sub-class of GRBs, called X-ray flashes (XRF), which emit the bulk of their energy around a few kiloelectronvolts, at energies significantly lower than classical GRBs. Apart from the softness of their emission the basic properties of XRFs have been found to be similar to those of classical ‘hard’ spectrum GRBs. This class of GRBs is analyzed in Chapter 4.
- The discovery of ‘dark bursts’ which have bright X-ray afterglows but no detectable optical afterglows. These dark bursts represent a significant fraction of BeppoSAX GRBs.
- The discovery of a correlation between the isotropic equivalent energy radiated by GRBs and the peak energy of their $\nu F_\nu$ spectrum. This correlation, and others discovered later, opened the possibility to use GRBs as cosmic rulers, as discussed in the final chapter (Chapter 9).

Clearly BeppoSAX was a watershed mission, which solved a significant part of the GRB mystery. With BeppoSAX it was not possible, however, to localize GRBs in seconds and to localize short GRBs. This would be the job of HETE-2 and particularly Swift, the following dedicated GRB missions that would reduce the time for GRB localizations to tens of seconds.

The beautiful results of Swift are reported in Chapter 4. Another mission, the small HETE-2 satellite dedicated to GRBs, had been launched before Swift. It has been quite successful because it was the first mission capable of distributing relatively precise GRB positions in several seconds, allowing the very early follow-up of the visible afterglow with robotic telescopes. As was to be the case for Swift, GRB data were analyzed on-board by powerful processors to determine the GRB position on the sky. In a few seconds this position was transmitted to the ground via a network of VHF receivers distributed along the Earth’s equator. Hence the scientific community was alerted within tens of seconds of the trigger. These fast alerts made it possible to uncover the complexity of the early optical afterglow.

The connection of long GRBs with supernovae was fully confirmed in March 2003, when HETE-2 detected an extremely bright GRB associated with a supernova at redshift $z = 0.17$. In July 2005, a HETE-2 alert led to the discovery of the first optical afterglow of a short GRB, which was soon followed (two weeks later) by a second observation triggered by Swift. These observations opened the window on the study of short GRBs, their energetics, their host galaxies, and their possible progenitors. Finally, the study of many X-ray flashes with HETE-2 showed that they are nothing other than soft GRBs. Thus, while confirming the link of long GRBs with supernovae, HETE-2 observations opened the field up by showing that short GRBs and XRFs share the basic properties of long GRBs, raising the question of what the progenitors of these events were. The models which have been proposed to explain the observed...
large diversity of GRBs (see Chapter 5) suggest that the emission of relativistic jets by dying stars, mergers, or magnetars might be a very general mechanism which would explain classical GRBs, XRFs, and short GRBs. Before analyzing the very ambitious Swift mission, Chapter 4 presents a remarkable contribution of the INTEGRAL ESA mission: the spectacular X-ray halos observed around GRB 031203 due to a large cloud of dust in our Galaxy located along the line of sight of the distant GRB. This is another illustration of the power of GRBs for the study of all the matter along the line of sight, from our Galaxy to the most distant host galaxies, through gigaparsecs of intergalactic medium.

The second part of Chapter 4 is entirely devoted to several major breakthroughs from the Swift satellite, launched in November 2004 and still in operation in 2008, with a typical rate of two GRBs detected per week (compared with two GRBs per month for BeppoSAX and HETE-2). The major advantage of Swift is its unique autonomous rapid slewing capability to track GRBs and their afterglows with minimum dead time. X-ray afterglows are observed as early as 50–70 s after the trigger and can be followed for days to weeks; this delay is only slightly longer for the Ultra-Violet/Optical Telescope (UVOT). The large effective area of its Burst Alert Telescope (5000 cm²) gives an excellent sensitivity and allows the detection of GRBs at higher redshifts, which explains why the mean redshift of Swift GRBs ($z \sim 2.75$) is more than twice as large as that of pre-Swift bursts.

Swift’s fast response has led to a completely new view of the early X-ray afterglow with well-defined phases which were previously unknown: after the prompt emission there is a phase of rapid decline which is followed by a late shallow phase preceding the standard decay, observed by BeppoSAX a few hours after the burst. In addition, about half the Swift GRBs exhibit late X-ray flares superimposed on the smooth decay. This newly discovered behavior is analyzed in Chapter 7, with proposed explanations. It seems in particular that the shallow phase and the X-ray flares suggest a long-lasting activity of the central engine (several hours). In Chapter 7 we also revisit the critical issue of the existence of achromatic breaks in the afterglows, which could be interpreted as jet breaks. The excellent coverage of X-ray afterglows with Swift led to the conclusion that truly achromatic breaks are rare, casting some doubt on the jet interpretation of the breaks seen in optical afterglows. This was another unexpected finding of Swift.

The report of Swift results would not be complete without the analysis of some exciting bursts that were detected by this satellite. We discuss four of them in detail. GRB 060218 was a low redshift X-ray flash ($z = 0.033$) whose supernova was studied in great detail, with the detection of the shock when the supernova broke out from the upper layers of the star. GRB 050904 at $z = 6.3$ is one of the most distant GRBs. It exploded when the Universe was $\sim$900 million years old, almost at the end of the epoch of re-ionization. Given the brightness of GRBs, these detections provide a unique way of probing the Universe when the first stars were formed. GRB 061007 is one of the brightest bursts detected by Swift, and it had excellent follow-up observations at all wavelengths. GRB 060614 is a nearby long GRB ($z = 0.125$) with no associated supernova, which raises questions on the diversity of GRB progenitors.
Finally, the discovery of several afterglows of short GRBs is a major Swift breakthrough. The initial sample, still limited, suggests that these bursts might have progenitors different from those of long GRBs. Even though the classification of long and short GRBs may be too simple, it nevertheless seems that the presence of two kinds of progenitors, corresponding to the deaths of massive stars and to the coalescence of compact objects, is required. These progenitors are extensively analyzed in Chapter 8.

The observational discoveries reported in Chapters 1–4 were accompanied by the development of GRB models explaining their main properties. Among them the GRB internal–external shock model has received a lot of attention and it rapidly became the “standard” model for GRB. In Chapters 5 and 6 we outline its main characteristics. In Chapter 5, we examine how internal shocks can explain the prompt emission starting from the behavior of relativistic fireballs. Since the photospheric emission of such fireballs cannot explain GRB non-thermal emission, shocks were rapidly added to the model: internal shocks and external shocks, which explain the prompt emission and the afterglow, respectively. We first present some basic elements of this standard model: the content of the fireball with its particles and magnetic fields, the shocks associated with ultra-relativistic fireballs, the acceleration of electrons in relativistic shocks and the resulting electron spectrum. The presence of these accelerated electrons and magnetic fields leads naturally to the consideration of the production of photons by synchrotron and/or synchrotron self-Compton emission. The characteristics of the resulting photon spectra, which can explain the GRBs and their afterglows, are briefly presented.

The second part of Chapter 5 concentrates on internal shocks. These shocks were introduced when it appeared that it was difficult to explain the fast temporal variability (at the level of tens of milliseconds) found in many GRBs with external shocks. Internal shocks can appear within a relativistic outflow if the Lorentz factor of the wind is variable. In this case, successive shells can have large relative velocities and fast shells injected after a slower one will eventually catch up and collide with it. This is the origin of internal shocks, assumed to be the result of multiple two-shell interactions. Internal shocks occur close to the central engine ($10^{14–15}$ cm) where the fireball is denser with higher magnetic field, producing synchrotron radiation in the X-ray and gamma-ray ranges. Many authors have studied the hydrodynamics of internal shocks and have shown that they can easily explain the multiple timescales observed in the GRB light-curves provided that the source itself is variable. In fact the observed temporal structure reflects the activity of the inner engine that drives the GRB. But internal shocks, even though they explain the properties of the GRB light-curves, meet with some difficulties. The two major issues are the compatibility of the spectral shape of the prompt emission spectra with the prediction of the synchrotron–shock model and the low radiative efficiency of internal shocks. We briefly mention the different solutions which have been proposed to resolve the first issue. The second point, concerning the efficiency of internal shocks, is more critical: the estimates of the ratio of the energy radiated in gamma-rays to the kinetic energy of the ejecta can sometimes reach 50%, while it is expected to be low, typically less than 10%. We analyze this question and the solutions which have been proposed to increase the radiative
efficiency of internal shocks. The debate on this question remains, and it has been amplified by the results of Swift, which suggest that late energy injection could increase the required radiative efficiency. Despite these open issues the internal shock model remains the favorite to explain the prompt GRB emission. Nevertheless, this chapter closes with a short discussion of other models, proposed to explain the prompt GRB emission, which avoid internal shocks. We mention in particular the electromagnetic model, with its Poynting flux, as a very attractive way to carry the energy in the flow. The possibility of a fireball which is more or less dominated by Poynting flux certainly needs to be studied thoroughly in the future.

The first part of Chapter 6 deals with the analysis of GRB afterglows in the framework of the standard model. It is followed by a second part which introduces the additional complexity needed to achieve a better agreement between the model and the properties of the afterglows observed with BeppoSAX and later HETE-2. We have indicated that internal shocks appear when the fireball coasts with constant Lorentz factor $\Gamma \sim 100$. This shell is cold, because the initial internal energy has been transformed into bulk kinetic energy. At the beginning of the coasting phase the medium has no influence on the expanding shell. As the shell progresses, it drives a shock into the interstellar medium (ISM) or the circum-burst medium; behind the shock the ISM is heated. As the shell radius $R$ increases, more ISM is shocked and the shell is progressively influenced by the ISM. This influence is significant when the energy of the heated ISM becomes comparable with the initial energy of the fireball. So the afterglow emission begins when most of the energy of the ejecta has been transferred to the shocked external medium.

The presence of an afterglow was anticipated when the fireball model was first proposed. It includes the forward and the reverse shock (which was not considered in the basic standard model). We indicate how the energy spectrum of the afterglow can be calculated assuming synchrotron emission and fast or slow cooling. The afterglow light-curves are also obtained for radiative and adiabatic evolution. The temporal decay index $\alpha$ and the spectral index $\beta$ of the afterglow are related by ‘closure relations’, which are formulated for various conditions: fast or slow cooling, adiabatic or radiative evolution, and various environments (ISM or wind). After the discovery of the afterglows of a dozen GRBs the gross characteristics of their light-curves and spectra have confirmed that the fireball model is robust. But with the observation of more and more afterglow light-curves at different wavelengths it rapidly became clear that the basic standard model was too simple. Indeed it considers spherical fireballs with an impulsive energy input and a single value of the Lorentz factor. In addition the highly relativistic expansion of the fireball is assumed to occur in a homogeneous external medium and in the adiabatic regime. The shock acceleration parameters $p, \varepsilon_e, \varepsilon_B$, are assumed constant for a given GRB. Finally, as we have said, the reverse shock is not taken into account. Even though this basic model was successful in explaining the first BeppoSAX results, it quickly became necessary to consider more complex conditions. These conditions are presented in the second part of Chapter 6.

The following elements have been progressively added to complete the basic fireball model:
Introduction

- Introduction of the reverse shock (within the ejecta), which can explain the strong optical flashes and radio flares.
- Collimation of the GRB outflow. This is a good way to avoid having exceedingly large radiated energy for some bright bursts at high redshift. The signature of collimation was searched for very early in optical afterglow light-curves and it is discussed in the context of the standard model. Another consequence of GRB beaming is the existence of ‘orphan afterglows’ from GRB jets which do not point towards the Earth. The search for such afterglows at X-ray, visible, and radio wavelengths is important to establish the presence of jets and to estimate the beaming factor without any knowledge of jet breaks. Finally, the energy distribution within the jet has been studied at length. Different types of jets have been proposed.
- The two most popular are the uniform jet and the universal structured jet. The main characteristics of these jet models are described, emphasizing the difficulty of inferring the jet structure based on the light-curve properties. In the future, polarization measurements of the afterglow may be helpful.
- The nature of the circum-burst medium is also a concern. The possibility of an inhomogeneous medium was considered very early on, through a dependence of density, $n$, on the distance from the source. The case of a wind ($n(R) \propto R^{-2}$) is particularly attractive because the progenitors of long-duration GRBs are assumed to be massive Wolf–Rayet stars. As these stars emit strong stellar winds, such an environment would be expected in the vicinity of long-duration GRBs. But, although a wind environment is preferred for some bursts, in a larger number of bursts a homogeneous circum-burst medium gives the best fits to the observations. Other possible properties of the circum-burst medium are analyzed, such as irregular density profiles, clumps of matter, and inhomogeneities due to interstellar turbulence. These characteristics are considered with their implications for the afterglow light-curves and the diversity of their temporal signatures.
- The introduction of variable energy injection is also discussed. Like the circum-burst medium, energy injection can contribute to produce more complex light-curves. Variable energy injection means that either the expanding ejecta is characterized by a distribution of Lorentz factors or that it consists of shells with similar Lorentz factors ejected at different times. This is the ‘refreshed shock’ scenario. Delayed energy injection is invoked to explain the shallow phase of the early X-ray afterglow and the X-ray flares discovered by Swift.

We close Chapter 6 with some examples that compare the models with some well-observed afterglows. This comparison centers on the presence of jets, the nature of the circum-burst medium, and the possible variability of the microphysical parameters, $\varepsilon_v, \varepsilon_B$. Finally the role of the inverse Compton mechanism, which is not considered in the basic standard model, is also mentioned. Even with these adjustments to the fireball model, many questions concerning the afterglow model were still
waiting for a solution in the pre-Swift era. The situation has become even more complex and possibly more critical after the beautiful results of Swift on the early afterglows. Discussing how models can explain these results is the main objective of Chapter 7.

The first part of Chapter 7 is devoted to the discovery of three features of early X-ray afterglows, which have had a decisive impact on the models: the steep decay phase after the prompt GRB emission, the shallow phase, and the X-ray flares. The steep decay phase is usually explained by the delayed prompt emission coming from the edge of the jet, the so-called curvature effect. But its success depends on the ‘right’ choice of the trigger time. The shallow phase has received a lot of attention and many possible explanations have been proposed. Among them are different jet models, prior activity, and time-dependence of the microphysical parameters. As all these models have difficulties, the most accepted one today is late energy injection. But this explanation exacerbates the issue of the efficiency of internal shocks. The late injection of energy implies that at the time of the internal shocks, only a fraction of the total kinetic energy is available and the efficiency of internal shocks must increase. These questions have given birth to new ideas, for example the explanation of the entire afterglow by a reverse shock, with a much reduced role of the forward shock, if any. So today the explanation of the shallow phase appears to be another open issue, in spite of large support from the community for the possibility of late energy injection. The third discovery is the X-ray flares. As for the shallow phase, the novelty of this phenomenon, which is present in about half of the GRBs, has led to many suggestions. The more accepted one involves late internal shocks due to the reactivation of the central engine at late times, the flares explained by internal energy dissipation taking place before the deceleration of the ejecta. Another explanation invokes a short-lived central engine which produces a tail of slower shells following the fast ejecta source of the GRB. Since the shallow phase and the X-ray flares seem to require a long-lasting activity of the central engine, we briefly explore how late activity of the central engine, erratic or continuous, can be obtained.

The second part of Chapter 7 examines the dearth of achromatic breaks now reported for a large number of Swift GRBs. Thanks to the XRT the search for X-ray breaks at the time of the optical breaks is routinely possible. Unfortunately it appears that truly achromatic breaks, simultaneously seen in the optical and X-ray domains, are more the exception than the rule. Moreover, when they are observed, their interpretation as jet breaks is often questionable. While jets must be present to avoid having GRBs with extreme energetics, there is a general feeling that it is difficult (at present) to use the afterglow to draw conclusions about the size of the jets and their geometry.

This analysis of these major Swift discoveries proves that Swift has already provided valuable information. These new data show, however, that our understanding of GRB afterglows in the pre-Swift era was clearly too simplistic. A lot of work and ingenuity will be necessary to preserve what we have considered until now to be robust models. But the exceptional data base that Swift is accumulating will help define the global statistical properties of GRBs which will finally be decisive for making progress in our understanding of these phenomena.
In Chapter 8 we present the long list of progenitors which have been proposed to explain GRBs. Two broad classes of models are considered to explain long and short GRBs respectively: the collapsar, which involves the explosion of a massive star, and the coalescence of two compact objects (NS + NS or BH + NS). Both kinds of systems lead to a black hole surrounded by a torus of matter, but the mass of the torus is much larger for collapsars, and the environment of these two systems is also different. An alternative to accretion as a source of energy involves pulsar-like activity, with rapidly rotating compact objects and very high magnetic fields, such as ‘millisecond magnetars’.

We first consider the collapsar model, based on the collapse of a massive WR star. In this model the explosion of the core of the massive star leaves a black hole and a torus of matter. A relativistic jet can be liberated along the rotational axis of the system and a GRB will be produced if the jet is able to pierce the star’s envelope. But this depends on the loading of the jet and the available energy. The energy that accelerates the jet can be extracted from various processes: neutrino annihilation which can tap the thermal energy released by viscous dissipation in the torus, the dissipation of energy by the intense magnetic fields generated by differential rotation in the torus, or the extraction of energy from the black hole itself by the Blandford–Znajek process. After various simulations of relativistic jet propagation and break-out in massive WR stars, the collapsar model seems to be secure. But there are many questions concerning the types of stars that can be the GRB progenitors. Several conditions have to be fulfilled: a black hole is needed, the envelope around the collapsing star has to be pre-ejected, and the collapsing star has to have an unusually large angular momentum. Because it is difficult for a progenitor to fulfill these three conditions it can be concluded that the nature of the stellar progenitor is not an entirely settled issue. In particular we do not know if the progenitors are single massive stars or massive stars in binary systems. The simulations also point out the crucial role of metallicity and of the magnetic fields. Low-metallicity massive stars in binary systems might be the best progenitors. A variant of the collapsar model is the supranova model, based on the explosion of a massive star which gives birth to a supra-massive neutron star with a delayed collapse to a black hole. Massive stars can also give birth to neutron stars with very large magnetic fields—typically $10^{15}$ G, like the magnetars—and very rapid rotation periods (one to few milliseconds). During the first seconds of their life these objects will slow down very quickly, emitting pulsar-like radiation with a huge luminosity. A few words are added on this class of models, in which the outflow is magnetically driven and Poynting-dominated. Observational evidence linking supernovae and long gamma-ray bursts is also reviewed.

For short hard bursts (lasting less than few seconds) the favorite progenitors are the merger of two neutron stars or of a neutron star with a black hole. As for the collapsar the evolution of such systems leads to the formation of a BH with a torus of matter, but a less massive one than in the case of the collapsar. We first consider the coalescence of two NSs. Since the torus is hot and very dense the energy can be extracted thanks to $\nu\bar{\nu}$ annihilation. The energy available for the relativistic jet is sufficient to power a short GRB if modest beaming is assumed. In addition magnetic processes can complete the energy extraction via magnetic fields threading the torus;
this mechanism can extract the rotational energy available in the torus. Finally, the Blandford–Znajek process may also take place, allowing part of the rotational energy of the black hole to be extracted. While there are many ways to power a relativistic jet, the problem of baryon contamination, which may seem less severe a priori, is nevertheless a concern even for coalescence models. Despite these uncertainties, NS + NS coalescences are today considered to be realistic progenitors of short GRBs.

Concerning BH + NS systems, the first calculations were pessimistic and it seemed difficult to have a disk around the BH. However, recent work based on simulations using fully relativistic hydrodynamics appear more favourable, and it seems that BH + NS coalescences can compete with models involving binary NS mergers. We complete the discussion with a few words about other possible mergers involving a white dwarf and a NS or a BH.

In this chapter we also address the issue of the galactocentric distribution of short GRBs, which provides a way to discriminate between the two types of progenitors, collapsars and mergers. It turns out that the type of host galaxy and the position of the GRB with respect to its host might be good tests to differentiate models. To close this chapter some recent results on short GRBs are presented, emphasizing the few discoveries of short GRB afterglows which seem to indicate that short GRBs and long GRBs have different progenitors and that short GRBs might be associated with the coalescence of compact objects. This conclusion still has to be considered with caution, given the small size of the short GRB sample.

In Chapter 9, before a short conclusion, we have chosen to address three topics that, in our opinion, complete the previous chapters. They are (i) the use of GRBs as cosmic rulers, as possible tracers of the first generation of stars, and as lighthouses illuminating the interstellar and intergalactic medium at high redshift, (ii) the expected signals from GRBs outside the electromagnetic spectrum, and (iii) some critical issues which remain linked to the fireball model.

The first point is the possible use of GRBs as cosmic rulers to complete and extend the present-day Hubble diagram obtained from observations of type Ia SNe. As GRBs have already been detected beyond \( z = 6 \), they can participate in the determination of cosmological parameters, since the observations of type Ia SNe are limited to redshifts lower than 2. But as GRBs are not standard candles a major concern is how the luminosity of GRBs can be standardized. Despite interesting suggestions we have not yet found a method to standardize GRB luminosities with sufficient accuracy to obtain new constraints on the cosmological parameters. This field will certainly continue to evolve in the coming years.

We also discuss how the high redshift of some GRBs and their extreme luminosity opens a new window on the exploration of the early Universe at the time of re-ionization when the first generations of stars appeared. The absorption lines in optical afterglows are very useful for measuring the metallicity, the dust content, and the kinematics of the interstellar medium at locations where star formation is taking place, i.e. the long GRB sites. It is also interesting to consider the rich domain of the host galaxies, which can be observed when the afterglow has vanished. The metallicity of the host galaxies can be studied with emission lines originating in the ionized warm gas of star-forming regions. This allows a comparison of the properties of GRB host
Introduction

galaxies with those of the average population of galaxies in the same range of redshifts. These topics are still in their infancy, but they have to be considered as major objectives of Swift and future GRB missions, beyond the pure understanding of GRBs and their physical models.

This leads to the second point addressed in this chapter, the signals which accompany or may accompany GRBs: the very high-energy photons and non-electromagnetic emissions like the Ultra High Energy Cosmic Rays (UHECRs), high-energy neutrinos and gravitational waves. High-energy gamma-rays have already been detected for several bursts thanks to Egret on CGRO. We show that the expectations are very high in this domain due to two space missions: GLAST/Fermi (launched by NASA in 2008) and the smaller AGILE (launched by ASI in 2007). Of course we also mention the ground observatories, which can track GRBs beyond hundreds of gigaelectronvols. They are very important for extending the observations of space missions, which cover the energy range 30 MeV–300 GeV. Thanks to these missions positive detections and even upper limits on high-energy and ultra-high-energy photon emission would allow us to complete our understanding of GRB models.

Concerning the non-electromagnetic signatures of GRBs, none of them has been detected yet, but their impact would be decisive. UHECRs may be produced in GRBs; this is expected but no decisive proof of their existence in GRB sites has been found. For high-energy neutrinos, thanks to the new generation of ground-based observatories, detection seems possible for nearby bursts and for the most optimistic models. They would provide good evidence that GRBs are able to produce UHECRs, even if other sources are possible, as shown by the recent HESS results. Gravitational waves might be associated with GRBs, especially the short ones, if they result from the coalescence of two compact objects. Of course the searches for gravitational waves are not at all limited to the study of GRBs, but their detection in association with a nearby GRB would make considerable progress possible in the identification of progenitors and especially those of short GRBs. Again in this domain the new generation of interferometers gives us great hope for a crucial discovery.

The third point of Chapter 9 is a short list of some of the major issues confronted by the most studied GRB model, which explains prompt and afterglow emissions mainly by internal and external shocks. The observations of Swift have raised significant questions, which have not all received satisfactory responses. Some authors have even questioned the possibility of producing the prompt GRB emission with shocks. Before going to such extremes we quickly review some important concerns with the fireball model. The list includes the origin and structure of the magnetic fields, the acceleration of particles in shocks, the values of the microphysical parameters, the density of the circum-burst medium, the origin of the breaks in the afterglow light-curves, the interpretation of the shallow phase observed in the early X-ray afterglows, the duration of the activity of the central engine, and so on. This short list of major issues is closed by considering some problems encountered by the progenitors of short and long GRBs: What are the massive stars which produce the long GRBs? Are short GRBs the result of the coalescence of two neutron stars or a black hole and a neutron star, or both, or something else?
All these serious concerns must not lead to an overly pessimistic conclusion. As we show in this book, the understanding of GRBs from the early beginning has progressed considerably; the models are not yet completely established but the most popular one has accumulated some beautiful successes. The association of long GRBs with supernovae has led to enormous progress in the identification of the progenitors of this type of burst. Of course neither the type of star—massive star, single or in binary system—nor its rotation rate, nor its metallicity, are known, but the framework seems solid. The shocks have been questioned, so more work has to be done on electromagnetic models, which transport energy in the GRB outflow by the magnetic field, whose dissipation produces the radiation without involving shocks. We conclude Chapter 9 with optimistic words about the future of this field and with a tribute to the pioneers who started the study of GRBs with small high-energy detectors on satellites and interplanetary probes, and who never gave up through the long dark years during which the clues on the nature of these mysterious GRBs were escaping.

Now we hope that the reader will be excited to read this GRB story, a modern scientific adventure, which was to start with the beginning of the space age about 40 years ago. This book should be taken as a detailed introduction to the field, to its fascinating history (Chapters 1–4), to the recent observational status (Chapters 4 and 7), and to our theoretical understanding of GRBs (Chapters 5 to 8). The field is now growing very fast and new applications of GRBs are appearing at a rapid pace. We hope that with this book the reader will be in a good position to understand the new developments that will enrich this field in the coming years.
1

The early times

1.1 INTRODUCTION

Gamma-ray bursts (GRBs) are short flashes of gamma-rays. They were fortuitously discovered by the VELA military satellites which were used by the USA to monitor possible nuclear explosions within and outside the atmosphere of the Earth. The first event was recorded in 1967 but the existence of gamma-ray flashes coming from the cosmos was not announced to the scientific community until 1973 (Klebesadel, Strong, & Olson 1973; Figure 1.1). GRBs are brief and bright, lasting from less than a second to tens of seconds (sometimes hundreds of seconds). It is impossible to predict when and from where they will arrive. During their short lifetime, they are so bright that they can outshine all other gamma-ray sources. This makes them easy to detect with small omnidirectional detectors outside the Earth’s atmosphere. Gamma-ray bursts were very quickly considered as an exciting new phenomenon by the astronomical community, and the rather easy way to detect them led many groups to explore this new field. In this book GRBs will be designated by the day of their detection, in this way GRB 790613 is the burst that has been detected on June 13, 1979. If more than one burst is detected in one day a letter is appended to the name: GRB 790613A is the first GRB detected on June 13, 1979, and GRB 790613B, the second.*

1.2 MORPHOLOGY AND TEMPORAL PROPERTIES OF GRBS

Very early it was realized that the light-curves of GRBs are quite variable from one burst to the next (Figure 1.2). The shapes are different as are the timescales which can vary from few milliseconds to hundreds of seconds. The time profiles may present

* Sometimes the first GRB is found on the ground, after the discovery of the second, and the letters do not reflect the time sequence.
multiple peaks or spikes, or they can be simple without any fine structure such as GRB 830801B (Figure 1.2). The shortest structures identified in the light-curves of long bursts last only a few milliseconds. For instance a 2 ms long spike was observed in GRB 790613 (Vedrenne 1981; Figure 1.2). This large variety of shapes has been tentatively used to classify GRBs, but without much success. However, the analysis of GRB durations has shown a bimodal distribution with two broad peaks around 0.3 s and 10–20 s (Figure 1.3; see also Hurley 1989a, Mazets et al. 1981b, Norris et al. 1984). Short GRBs are not only shorter, they also have a harder spectrum (Dezalay et al. 1992, Kouveliotou et al. 1993). Another property was also noticed quite early: at high energies GRB light-curves are sharper with shorter duration (Figure 1.4). This
Figure 1.2. Light-curves of GRBs: some examples of their diversity. The upper panel (a) shows the light-curves of GRB 790613, GRB780918, and GRB 820313A detected by the detectors of the Franco-Soviet SIGNE collaboration on the Venera Venus probes (Hurley 1989a). The lower panel (b) displays a GRB without fine time structure, GRB 830801B (Kuznetsov et al. 1986).
effect will be later quantified with BATSE (the Burst and Transient Source Experiment on the Gamma-Ray Observatory—GRO).

1.2.1 Periods

Although the majority of GRBs have multiple peaks, searches for periodicities in their light curves have generally been unsuccessful, even if the evidence for periods at the level of few seconds were reported. For instance Wood et al. (1981) indicated a 4.2-s period for the GRB 771029 and Kouveliotou (1988) a 2.2-s period lasting for seven cycles for the most intense GRB observed by SMM, the Solar Maximum Mission, GRB 840805. Seven events in the Konus catalogue were characterized by an approximate 5 s delay between the outbursts in a multipeak GRB (Mazets & Golenetskii 1981b). Similar characteristic time delays (5.75 s) were observed in the 1979 January 13 event between pulse pairs which were very similar (Barat et al. 1984a). Eventually Schaefer and Desai (1988) showed that the only indisputable periodicity was the 8-s period observed in the 1979 March 5 event during 22 cycles. But we will see that this GRB is not typical and was quickly suspected of belonging to another class of burst (Cline et al. 1980, Cline et al. 1982), later called soft gamma-ray repeaters (SGRs).

1.2.2 Soft X-ray emission

Few observations were made at energies below 10 keV, we will see in the next paragraph that the energy content below 10 keV is a few percent of the total. This
Figure 1.4. Narrowing of the width of the peaks with energy in GRB 830801B (Kuznetsov et al. 1986).
soft X-ray emission has been detected by Ginga between 2 keV and 10 keV (Yoshida et al. 1989, Murakami et al. 1991, Strohmayer et al. 1998), but also by Vela, P 78-1 and Hakucho satellites (Terrel et al. 1982, 1983, Laros et al. 1984, Katoh et al. 1984). The soft X-ray emission can appear a few seconds before the onset of the GRB and it can last a few tens of seconds after the end of the GRB.

In the 1980s the most evident conclusion concerning the morphology of GRB light-curves was their great diversity; in spite of some fortuitous resemblances, it can be said that each GRB seems to have its own signature. Let us see now if the same diversity exists for their energy spectra.

1.3 SPECTRA OF GRBS

1.3.1 The continuum

When GRBs were discovered, the detectors on the Vela spacecraft were not equipped to do the spectroscopy of the gamma-ray events but very quickly Cline et al. (1973) gave the first spectra obtained with a semi-omnidirectional X-ray detector on-board IMP-6. The differential spectra of six GRBs appeared to be well represented by an exponential of the form: \( dn/dE = I_0 \exp(-E/E_0) \) photon/cm\(^2\)/keV/s with a characteristic energy \( E_0 \) of 150 keV.

Quickly, other results were reported, for instance the \( \gamma \)-ray and X-ray spectrometers on Apollo XVI observed a burst (GRB 720427) which was also seen by Vela 6A. The spectrum of this burst could be fitted by two power laws: from 2 keV to 200 keV with an exponent \( \alpha \) of \(-1.4\), and from 0.3 to 3 MeV with an exponent of \(-2.6\). An exponential response could also be used from 70 keV to 1 MeV but it was unsatisfactory beyond 1 MeV (Metzger et al. 1974). The same year Imhof et al. (1974) reported the observation of a GRB (GRB 721218) with a 50-cm\(^3\) Ge detector on-board a low altitude polar orbit satellite called 1972-076B. The spectra obtained in six different short accumulation periods could be fitted with both a power law or an exponential function. In 1975, Cline and Desai reported new observations of GRBs from IMP-7. Nine events were spectrally observed (Figure 1.5). They were not only similar but were actually consistent with a common fit consisting of a 150 keV exponential function from 100 keV to about 400 keV, tangent to a power law of index \(-2.5\) above that energy (dashed curves). Two power laws of index \(-1\) below 100 keV and \(-2.5\) above several hundred kiloelectronvolts could fit also the spectra (Cline & Desai 1975).

It has to be added that when access to energies of a few kiloelectronvolts was possible, with instruments such as the X-ray spectrometer on OSO 5 or the solar X-ray instruments on OSO 5 and OSO 7, GRBs were detected down to these energies (Kane & Anderson 1976, Wheaton et al. 1973, Kane & Share 1977). Nevertheless below 10 keV the photon number spectra often flattened or even turned over steeply (Katoh et al. 1984). This characteristic was called the X-ray paucity, because only a few percent of the total energy emitted in a GRB appeared in the X-ray domain.
Figure 1.5. Spectra of nine GRBs observed with IMP-7. An exponential spectrum with a power-law tail at high energies, which gives reasonable fits, is shown for each event by a dashed curve (Cline & Desai 1975).

On the USSR side, a very significant contribution was made by the Mazets group from A. F. Ioffe Physical Technical Institute. This group started to work on gamma-ray bursts at the beginning of the 1970s using various satellites and probes, first the Kosmos 461 and Meteor satellites. But their decisive contribution was obtained with the Konus experiments on the Venera probes (Venera 11, 12, 13, 14). With six detectors on each probe this scientific equipment was certainly one of the first and the most complete to be specifically designed for GRB studies: the detectors measured the time profiles and the spectra and provided crude localizations—thanks to the six scintillation counters pointing in directions perpendicular to the six faces of a die, and covering the whole sky. The relative response of the different detectors with respect to the arrival of a burst of photons allowed obtaining crude localizations at the level of few degrees (Mazets & Golenetskii 1981a,b). From September 1978 to February 1980, the spectra of 143 bursts were measured between 30 keV and 2 MeV. The conclusions concerning the spectra were the following (Mazets et al. 1981b): although no simple, uniform law fitted all the spectra, to a rough approximation many of them
followed a power law: \( dN \sim E^{-\alpha} dE \) with \( \alpha \) varying from 1.3 to 2.5, the most usual values being in the range 1.5–1.7. As already noticed by T. Cline some spectra were better described by an exponential law \( dN \sim \exp(-E/E_0) dE \) with \( E_0 \sim 100–200 \text{ keV} \). But for most GRBs the best fit of the spectra was obtained with another law: \( dN \sim E^{-1} \exp(-E/kT) dE \), typical of the thermal bremsstrahlung emission of a hot optically thin plasma (OTTB), with a temperature \( T = (0.5–20) \times 10^9 \text{ K} \) (corresponding to 50–2000 keV). This type of fit was also reported for the well studied GRB 720427 (Gilman et al. 1980; Figure 1.6).

So, even if there was not such a large diversity in the spectra, compared with the time profiles of GRBs, the quite similar average spectra observed by IMP-6 and IMP-7 were not confirmed for all the GRBs. More important, the fits depended on the energy range covered by the different instruments. Fenimore et al. (1982a) noted that

![Figure 1.6](image-url)

**Figure 1.6.** An example of a GRB spectrum (GRB 720427) obtained with the X-ray and gamma-ray spectrometers on Apollo 16. A thermal bremsstrahlung fit with \( kT = 500 \text{ keV} \) is drawn through the X-ray and gamma-ray data (Gilman et al. 1980).
the fit of GRB spectra by optically thin thermal bremsstrahlung demanded high
temperatures \( (kT \sim 300\text{ keV}) \). At such high temperatures the free–free cross-section is
much smaller than the Compton cross-section. The Compton scattering should
thus dominate the spectral formation process, leading the authors to investigate
emission mechanisms based on Compton scattering. They considered black-body
emission undergoing inverse Comptonization by a much hotter overlying plasma.
The black-body was an X-ray source with \( kT_{BB} \sim 2.4\text{ keV} \), and the overlying plasma
had a temperature \( T_e \) with \( kT_e \sim 150\text{ keV} \). Their conclusion was that many GRB
spectra could be consistent with an X-ray source undergoing inverse comptonization
by a hot overlying plasma. This point concerning the fit of the continuum is impor-
tant because, as we will see below, the presence of lines in the energy spectra might be
required or not depending on the choice of the mechanism (OTTB vs comptonized
black body) producing the continuum (Fenimore et al. 1982b).

Another property of GRB’s spectra is their fast temporal dependence, which was
discovered very early. Fast spectral variability was reported for the bright burst GRB
781119 (Barat et al. 1981, Mazets & Golenetskii 1981b, Vedrenne 1981; Figure 1.7).
Very often, the spectrum was observed to soften during the evolution of the burst.
The gamma-ray spectrometer (GRS) on-board SMM illustrated this kind of evolution,
quantified by a hardness ratio (Figure 1.8; Norris et al. 1986). At lower energy,
where a thermal bremsstrahlung function could fit the GRB spectra, Golenetskii et al.
(1983) demonstrated the existence of a good correlation between the temperature and
the luminosity in many bursts. However, these variability laws could not be con-
sidered as universal. Kargatis et al. (1994) considered the continuum spectral evolu-
tion of 16 GRBs detected by the Franco-Soviet SIGNE experiment using different
fits, with a simple thermal bremsstrahlung and with synchrotron models. They found
no single characteristic of spectral evolution. Correlations between instantaneous
burst intensity and spectral temperature have been found for seven GRBs with a
much larger range of correlation indices than that suggested by Golenetskii et al.
(1983): \( L \propto T^\alpha \) with \( \alpha = 2.2 \pm 1 \) (Kargatis et al. 1994). Finally, this fast variability
makes difficult using time-averaged GRB spectra to identify the radiation processes
producing the gamma-rays. Fitting average GRB spectra with known processes, like
OTTB, comptonized black-body, or synchrotron emission is often meaningless, given
the amount and timescale of the spectral variability.

Independently of the precise shape of the spectra, a general property of GRBs is
the non-thermal nature of their emission, with the maximum of their energy emitted
in the hard X-ray or soft gamma-ray domain (100 keV to 1 MeV, see Figure 1.9). The
paucity of X-rays is another general property of GRBs.

Various thermal and non-thermal radiation processes have been considered to fit
the continuum of the GRB spectra obtained with a large diversity of instruments.
They are analysed by Lamb (1984a,b). In the 1980s the comparison of these possible
processes with the observations seemed to rule out several emission mechanisms,
including bremsstrahlung and inverse Compton scattering from a shock. They
appeared consistent with inverse Compton scattering off a thermal distribution of
hot electrons, with thermal synchrotron emission, and possibly with synchrotron
emission from a power law distribution of ultrarelativistic electrons. Later, Lamb
Figure 1.7. Spectral evolution in a well-structured GRB 781119 observed with the SIGNE experiments onboard Venera 12 and 11 (Vedrenne 1981). The lower panel (b) shows the six spectra corresponding to intervals A–F in the light-curve.
Figure 1.8. Spectral evolution of pulse structures in GRB 821104 observed with Solar Maximum Mission satellite (SMM). The evolution of the power-law spectral index obtained between 144 and 440 keV and of the hardness ratio indicate a clear hard-to-soft evolution for the different pulses inside the GRB (Norris et al. 1986).

(1988) came back on the two classes of models (thermal and non-thermal). Thermal models were excluded by the observation of power law spectra extending to megar electronvolt energies with SMM, and various non-thermal models were suggested.

Repeated Compton scattering of soft photons by non-thermal electrons have been proposed by Zdziarski and Lamb (1986). Synchrotron emission from non-thermal power law electron distributions could also fit the power law emission at energies $E > 1$ MeV but produced X-rays well in excess of the observed flat X-ray spectrum. So, Brainerd and Lamb (1987) proposed a two component synchrotron model. Ruderman and Cheng (1988) suggested that latent rapidly spinning pulsars reignited by weak accretion of interstellar matter could radiate with the intensity and spectrum of most hard GRB sources (see also Sturrock (1986) and Zdziarski (1988).
for other models). Higdon and Lingenfelter (1990) and references therein can be consulted for a more complete review on this topic. These fits of GRB spectra were complicated by the presence of cyclotron features discussed in the next paragraph.

1.3.2 Spectral lines

Besides the continuum, cyclotron and annihilation lines were reported for many bursts (Mazets et al. 1981a; see Figure 1.10). In fact more than 20% of the GRBs seemed to exhibit spectral features resembling absorption and emission lines. The absorption lines were observed in the 30–70 keV energy range always in the initial phase of the burst. The absorption lines would be due to magnetic bremsstrahlung absorption at the cyclotron frequency \( \omega_H = eB/\pi c \), the energy observed for the absorption features leading to field strength \( B \sim (2–6) \times 10^{13} \) gauss. These features were present during limited fractions of the light-curve of the GRB. They were naturally interpreted as a convincing proof that the sources might be neutron stars (NS) with high magnetic fields. The emission lines, reported around 400 keV, could be annihilation lines (at 511 keV), with an energy lowered by the strong gravitational field at the surface of the NS.

Other instruments observed such features: Huet et al. (1984) reported an absorption feature at 55 ± 5 keV with a width of 13 ± 3 keV in GRB 780325 observed

![Figure 1.9. A typical GRB spectrum illustrating the non-thermal nature of the emission and the peak power evolution during the burst: GRB 830801B observed by the SIGNE Prognoz 9 experiment (Kuznetsov et al. 1986).](image-url)
with the High Energy X-ray and Low Energy Gamma Ray Experiment on HEAO-1. In the bright GRB 800419, Dennis et al. (1982) reported the presence of a broad absorption feature, extending down to 28 keV, and varying on a timescale of \( \lesssim 0.5 \) s. This observation was made with the hard X-ray burst spectrometer on SMM. For GRB 781119, ISEE 3 using a germanium spectrometer reported the observation of two spectral features: a broad line at 420 keV and a narrower one at 740 keV (their figure 6). That line was attributed to \(^{56}\text{Fe}\) nuclear de-excitation at 847 keV that was gravitationally redshifted by 10–18\% again compatible with current NS models (Teegarden & Cline 1981). The Franco-Soviet SIGNE experiments on the Venera 11 and 12 probes observed a spectral feature around 400 keV for GRB 781104 (Barat et al. 1984c). But is is clear that the Konus instruments have provided the most important and homogeneous population of GRBs presenting features around 400 keV in emission and 30–70 keV in absorption.
Nevertheless the reality of these features raised a lot of debates. A line seen by one group was not always confirmed by another observing the same burst with a different instrument. For GRB 811231, for example, the gamma-ray spectrometer of SMM exhibited no evidence for the presence of an emission feature reported by Konus between 400 and 500 keV (Nolan et al. 1984). Fenimore et al. (1982a) disputed the possible features around 400 keV detected by Mazets and his collaborators in GRB 781104 and GRB 781119, and in other bursts. They claimed that the hardness of the continuum model might affect the presence of such lines. Specifically, the assumption of a wrong continuum shape combined with gain shifts could introduce spurious line features (Fenimore et al. 1982b). In addition, due to the fast variability of the continuum (for instance 19 November 1978, Figure 1.7), a spectrum integrated on a time scale significantly larger than the variability of the burst might also mimic the presence of a line or a bump (Vedrenne 1991). Another possibility to produce artifacts might be the poor knowledge of the instrument gain. In fact the low energy part of the spectrum between 20–60 keV is particularly sensitive to small changes in the gain or the response function. This might cast some doubts on the photon spectra observed in this energy range.

The effect of the assumed continuum was studied for GRB 800419 for which the Konus experiment observed cyclotron lines while ISEE-3 reported no line. If a soft spectrum is assumed, such as optically thin thermal bremsstrahlung, an absorption feature might appear, whereas a harder spectrum, such as a comptonized black body, does not require the feature (Fenimore et al. 1982b; Fenimore, Klebesadel, and Laros 1983; see their figure 1). They also stressed that for NaI and CsI scintillators the dead layer may represent a crucial issue. This dead layer is a several micrometer thick layer on top of the crystal, which does not scintillate because the manufacturing of the crystal destroys the crystalline lattice in this zone. This dead layer crucially modifies the spectral response of the detector from a few to a few tens of kiloelectronvolts, its impact on the analysis of GRB spectra below 60 keV has been studied by Niel, Jourdain, and Roques (1990). This problem can be avoided by using cleaved crystals as in HETE-Fregate (see Chapter 4).

Finally, doubts on the presence of lines can also arise from the large number of GRB spectra which are treated and for a given burst the number of spectra which are analyzed with different timescales. The large number of GRB spectra searched for leads to the ‘discovery’ of a few spectra displaying interesting features; however, these features can also be explained by statistical fluctuations.

By the 1990s, if doubts still existed concerning the reality of the 400 keV emission lines, the absorption features interpreted as cyclotron absorption lines were completely accepted after their observation in three GRBs by the Japanese satellite Ginga (Figure 1.11). The lines at ~20 and ~40 keV seen in the spectrum of GRB 880205 (Murakami et al. 1988, Fenimore et al. 1988) have been interpreted as photon scattering resulting in transitions from the ground level to the first and second Landau levels in a strong magnetic field ~1.7 × 10^{12} G (Fenimore et al. 1988, Lamb et al. 1989, Wang et al. 1988, 1989). In this interpretation, the 20 keV line resulted from cyclotron resonant scattering of photons with energy near the first harmonic, and the 40 keV line resulted from Raman scattering of photons with energies near the
second harmonic. Yoshida et al. (1991) reported another detection of cyclotron lines seen in GRB 890929. The characteristics of the three bursts with cyclotron lines are given in their Table 1. For this last burst the energies of the lines are higher than in the first two GRBs with cyclotron lines (see Figure 1.11). It is clear that the confirmation by Ginga of the presence of cyclotron lines in the spectra of GRBs, essentially at the beginning of the burst (in spite of some doubts which have been expressed) reinforced strongly the idea of associating GRBs with neutron stars with a large magnetic field. Another observation of Ginga reinforced the NS hypothesis: the detection in some GRBs of soft X-ray tails which could be fit with black-body spectra at a temperature of a few keV (Murakami et al. 1991). It was possible to estimate a source radius of: \( R \sim (0.7 \, d/1 \, \text{kpc}) \) km. For GRBs in the galactic disk the values of \( R \) were consistent with the typical dimensions of neutron star polar cap regions.

GRB spectra, mainly the lines, provided observational evidence supporting the emission of these events in polar cap regions of strongly magnetized neutron stars.
(a generic GRB spectrum is given in Figure 1.12 with all the possible features reported in the literature, from Liang (1989); see also Figure 1.21). This model was also backed up by the very short time structures (at the level of several milliseconds) observed in the light-curves, which led to infer small radiating regions with typical sizes of a few hundred kilometers. A good way to confirm these views was to look at the GRB angular distribution of localized GRBs, which is analyzed in Section 1.4.

### 1.4 ANGULAR DISTRIBUTION

Measuring the distribution of GRBs on the sky was a good way to approach their origin. Galactic sources in particular should be concentrated in the galactic plane. Omnidirectional gamma-ray detectors were generally used for GRB detection. While a single detector is unable to give the arrival direction of the burst by itself, a crude direction (at the level of a few degrees) can be obtained by comparing the count rates received by different detectors on the same spacecraft if they have a non-isotropic angular response and different orientations. Simple, flat detectors were used because the response of a detector to a parallel photon flux varies approximately as the cosine of the angle of the photon beam to the normal of the detector. Mazets and collaborators were the first to use this method with the Konus experiment (Mazets & Golenetskii 1981a). This technique would also be used later by BATSE on the Compton Gamma-Ray Observatory (CGRO).

Another technique, illustrated in Figure 1.13, can be used to localize GRBs if at least three satellites or probes detect the same burst. The difference $\Delta t$ of arrival times at two satellites allows definition of a cone with an opening angle $\theta$ (determined by $\cos \theta = c \Delta t/D_{12}$, where $D_{12}$ is the distance between the two satellites). With three
Figure 1.13. The triangulation method which has been largely used to localize GRBs. This method is still used in the InterPlanetary Network (IPN). The precision of the localization increases with the distance between the satellites, as shown in panel C (bottom) (Vedrenne 1981).
satellites, two cones are obtained; their two intersections give the two possible directions for the GRB. This triangulation method has led to the localization of many bursts through the Interplanetary Network (IPN).

Konus and the IPN, with different methods and for different GRBs, obtained maps of nearly one hundred GRB positions on the sky. In fact rapidly one catalog and two maps were available for an analysis of the angular distribution of GRBs (Mazets et al. 1981b, Klebesadel et al. 1982, Atteia et al. 1987; Figure 1.1.4). The analysis of the distribution of 88 GRBs localized by the IPN showed that this distribution was isotropic within the statistical limits (Atteia et al. 1987). In the same way the map obtained by Mazets et al. (1981b) did not show any concentration of GRBs in the galactic plane which could indicate a disk population for the sources. The same can be said with respect to the Galactic Center or any prominent extragalactic regions such as the Virgo Cluster, Andromeda Nebula M31 or the Magellanic Clouds (Hurley 1989b, Hartman 1991). This last result is important because the absence of correlation with galaxies or nearby clusters of galaxies rejects extragalactic GRBs to distances larger than 100 Mpc (the Virgo Cluster is at about 20 Mpc). This implied a typical luminosity of \( \sim 10^{30} \text{ergs s}^{-1} \) for extra-galactic GRBs, while it is only \( 10^{38} \text{ergs s}^{-1} \) if the GRBs were associated with galactic neutron stars typically at 100 pc. This huge difference in the involved energies is an important reason which explains why, in the 1980s, most theoreticians in the field tried to explain the GRBs by a galactic NS population, in the disk of the galaxy or in an extended galactic halo.

To quantify the isotropy of GRBs, one can naturally count the GRBs in two hemispheres one towards and one away from the Galactic Center. But this approach does not utilize all the available positional information: single error boxes but also double error boxes or error rings defined by the triangulation method. In order to be more precise Hartman and Epstein (1989) analyzed the angular distribution of localized GRBs by calculating the low-order multipole moments of their distribution on the celestial sphere and by comparing them with Monte Carlo simulations of an isotropic parent angular distribution including the same type of localizations or selection effects. The dipole moment \( \cos \theta \) (\( \theta \) is the angle between the object and the Galactic Center) measures the degree of concentration of objects in the sky toward the Galactic Center. The quadrupole moment \( \langle \sin^2 b - 1/3 \rangle \) (\( b \) is the galactic latitude of the object) measures the degree of concentration of objects toward the Galactic Plane. If the distribution of the objects is isotropic and homogeneous \( \langle \cos \theta \rangle = 0 \) and \( \langle \sin^2 b \rangle = 1/3 \). On the other hand, objects concentrated towards the Galactic Center would have \( \langle \cos \theta \rangle \) close to 1 and objects concentrated in the galactic plane would have \( \langle \sin^2 b \rangle \) close to 0. Another indicator, called \( \langle V/V_{\text{max}} \rangle \), can be used to get a measure of the distribution of the sources in the radial direction (see Section 1.5). The combination of these quantities allows testing the distribution of GRBs in three dimensions.

Measuring the dipole and quadrupole moments of GRBs in the catalog of Atteia et al. (1987), Hartman and Epstein (1989) demonstrated the isotropy of the angular distribution of the sources included in this catalog. In addition to this analysis Hartman and Blumenthal (1989) studied the angular clustering properties of GRBs taken from the two catalogs (Golenetskii 1988, Atteia et al. 1987). These properties
Figure 1.14.
Maps of GRBs localized by KONUS experiments (Mazets et al. 1981b, panel a) and by the IPN (Atteia et al. 1987, panel b). The two distributions are merged in the bottom panel (c) which contains 171 GRBs (Vedrenne 1991).
were compared with those of known extragalactic populations such as galaxies or
galaxy clusters. They used for this purpose the two-point angular autocovariance
function \( w(\theta) \) which measures the deviation from a random distribution on the sky.
The correlation of both data sets were consistent with \( w(\theta) = 0 \) over two orders of
magnitude in angle \( (\theta \sim 1–100^\circ) \). Moreover, this angular covariance function \( w(\theta) \)
has been used by the authors to conclude that if GRBs were associated with galaxies
the survey already made was complete out to 140 Mpc at least. This corresponds to a
total gamma-ray burst energy of \( \sim10^{50} \) erg for a typical distance of 100 Mpc and
GRBs with a typical fluence of \( 10^{-4} \) erg cm\(^{-2} \) (the fluence is the flux received from the
GRB integrated over the duration of the burst). This is much larger than the energy
emitted in \( \gamma \)-rays by a typical type II supernova (Pinto & Woosley 1988) whose energy
is radiated over \(~400\) days, and not in a few seconds! To appreciate the amount of
energy which would be involved in extragalactic GRBs, another interesting compar-
ison can be made with the \( \gamma \)-ray luminosities of active galaxies, which are typically of
the order of \( L_{\gamma} \sim 10^{46} \) erg s\(^{-1} \).

For galactic models in which the sources are distributed in a disk with a scale
height \( H \), the limits on \( w(\theta) \) were also used by the authors to conclude that the
survey depth \( D \) of the Konus catalog should be less than two times the scale height of
the disk: \( D \leq 2H \). Larger \( D/H \) values would have led to strong positional auto-
correlation on the sky which was not observed.

In conclusion, these studies demonstrated that the two main GRB catalogs
already available in the 1980s (Mazets et al. and Atteia et al.) indicated clearly
and quantitatively that GRBs had an isotropic distribution and that no angular
clustering could be observed. The analysis of the angular distribution also gave
the opportunity to estimate a lower limit for the repetition timescale of GRBs which
could vary from 0.5 to 10 years depending on whether the bursts were assumed to be
mono-luminous or to have an extended luminosity function (Schaefer & Cline 1985,
Atteia et al. 1987). Finally, these studies allow us to appreciate the volume of
information which was available by the end of 1980s and the importance of the
conclusions which could be already given at the time of the launch of GRO with
BATSE. To complete this analysis we will look at another property of the GRB
population, their intensity distribution.

### 1.5 INTENSITY DISTRIBUTION

For galactic sources, the isotropy was puzzling. In fact all other known populations
of galactic neutron stars were more or less strongly concentrated in the galactic plane
or in the galactic bulge. The most accepted interpretation of isotropy was that GRB
sources were nearby galactic objects and that the instruments were not sensitive
enough to see them beyond the scale height of the disk of the Galaxy. It was thought
that more sensitive detectors (seeing more distant sources) would be needed to
disclose the galactic structure of the GRB population. Since the IPN and Konus
localized only half the GRBs they detected, the brightest ones, the work based on
localized GRBs did not allow checking of the distribution of the full sample of
detected GRBs. A method to use the full sample was to check the spatial homogeneity of localized and non-localized sources. This can be done by measuring the intensity distribution of GRBs. For local, spatially homogeneous objects the number of visible sources increases as the cube of the distance, while the brightness of the sources decreases as the square of the distance. Consequently for a homogeneous distribution of standard candles, the number of visible sources must grow like the intensity at the power $-1.5$. This is how the intensity distribution provides insight into the spatial homogeneity of the source distribution. This fact was realized early (Strong & Klebesadel 1974, Fabian 1975).

The first measure of the GRB intensity used for this purpose was the total fluence $S$, i.e. the flux integrated over the whole duration of the burst. As we have just indicated if the sources are uniformly distributed in space the frequency of observations of bursts with fluxes or fluences greater than some value $S$: $(N(> S))$ must be proportional to the $-3/2$ power of that flux or fluence. A deviation of the observed size frequency distribution (log $N$–log $S$) from a $-3/2$ power law indicates a non-uniformity in the spatial distribution of the sources. For instance, for sources homogeneously distributed in the galactic disk with a scale height $h$, if the accessible volume is a sphere of radius $R$ smaller than $h$, $N(> S)$ is proportional to $S^{-3/2}$. If the sources can be observed beyond this distance $h$ there will be a flattening of the log $N$–log $S$ curve, with its slope smoothly changing from $-3/2$ to $-1$, characteristic of sources distributed in a plane. Therefore if the log $N$–log $S$ curve has a slope $-3/2$ for large values of $S$ (nearby sources) and presents a flattening at low $S$ values, it must indicate a disk distribution of sources and therefore it should support a galactic origin of the detected GRB population. Figure 1.15 illustrates these different situations.

Figure 1.16 summarizes the different observations of the number of GRBs $N(> S)$ vs $S$; this figure includes satellite observations above fluences $S \sim 10^{-5}$ erg cm$^{-2}$ and balloon observations obtained with more sensitive detectors with larger collecting area, allowing to reach fluences between $10^{-8}$ and $10^{-7}$ erg cm$^{-2}$. On this figure IMP-6, IMP-7, and Konus results seem to indicate a flattening of the log $N$–log $S$ curve below $3 \times 10^{-5}$ erg cm$^{-2}$, and balloon flights seem to confirm this flattening. But many discussions about the significance of these balloon observations and their intercomparison with the satellite observations took place. The measure of Carter et al. (1976), for instance, was disputed by Cline and Schmidt (1977) who determined a new limit, higher by two orders of magnitude. This new upper limit was no more in conflict with the $S^{-3/2}$ law extrapolated to low fluences (between $10^{-7}$ and $10^{-6}$ erg cm$^{-2}$) from the satellite measurements obtained above a few $10^{-5}$ erg cm$^{-2}$. Searches for GRBs using balloon flights with large plastic scintillators (8,900 cm$^2$ in area, 5 cm thick) contemporaneous with satellites did not allow the detection of coincident GRBs. The upper limit around $10^{-6}$ erg cm$^{-2}$ which was obtained was compatible with an extrapolation of the size spectrum to low fluences with a power law of index 1.5 (Cline et al. 1977). Herzo et al. (1976) reported another balloon observation with a Compton telescope. They detected one burst on May 14, 1975. The corresponding point they obtained on the log $N$–log $S$ curve allowed the conclusion that this observation was also consistent with the $S^{-3/2}$ law. The conclusion was about the same after a balloon flight with a large volume
Figure 1.15. This is the schematic illustration done by Briggs (1995) of the model constraints imposed by observing the 2D angular distribution and the intensity distribution of GRBs. The top panel shows a cross-section through a postulated exponential disk population of GRB sources. The dashed line is the Galactic plane and the star ∗ the solar system. The circles indicate the spherical volumes in which sources are detected. $R_S$ corresponds to the accessible volume for an instrument of poor sensitivity, $R_D$ is the volume which can be observed by another instrument with better sensitivity. The lower panels show the angular and intensity distributions of the GRBs detected by the two instruments. On the left the location of GRBs is given in galactic coordinates. The anisotropy of the distribution invisible for the RS case begins to be detected in the second case, the instrument being observing beyond the scale height of the sources distributed along the Galactic plane. This difference is also visible in the intensity distribution on the right ($\log N$–$\log P$ curves). In the $R_D$ case the distribution has a deficit for the low $P$ because the instrument is able to detect sources in regions where there are few sources, as can be seen in the top panel.

NaI(Tl) scintillator, 13 in. in diameter by 6 in. thick in 1976 (Johnson, Kurfess, & Bleach 1976).

Nevertheless other balloon experiments (White et al. 1978, Fishman et al. 1978, Bewick et al. 1975) seemed to indicate a real deficit of bursts below
Figure 1.16. The log $N$–log $S$ distribution which is the number of bursts per year $N(>S)$ with fluence greater than $S$ (erg cm$^{-2}$) as a function of $S$. The theoretical curves and their parameters are from Jenning and White (1980) and correspond to disk models with various values of distribution scale height $\beta$ (in parsecs) and with different burst rate densities given by $n(\rho) = n_0 \exp(\rho/\rho_0)$; where $\rho$ is the galactocentric radius vector. The size of the central concentration $\rho_0$ is in kiloparsecs. This model was generalized to include an intrinsic burst luminosity distribution (Jennings 1982). The references of the experimental results are given in Jenning and White (1980). The interpretation of this figure is discussed in the text (from Vedrenne 1981).

$S \sim 10^{-6}$ erg cm$^{-2}$. Fishman et al. (1978) flew on a balloon large NaI scintillation crystal arrays, a prototype of the Burst and Transient Source Experiment (BATSE) which were to be flown on NASA’s Gamma Ray Observatory (GRO). The detectors had a response similar to the Konus experiment, but nearly 100 times its effective area. During two balloon flights (64 hours of cumulated observation) about 40 GRBs
were expected if they were homogeneously distributed in space (i.e. following the $S^{-3/2}$ law); only one was observed. Fishman et al. concluded that there was a true dearth of faint sources. However, many factors seemed to render a balloon-borne search less efficient than it may have been intended to be, and Cline et al. (1977) raised some doubts on the reported deficit of low-fluence GRBs.

Surveys of fast X-ray transients can also be used to set bounds on the faint end of the log $N$–log $S$ curve. Among these fast transients, fewer than half of them have been identified with active coronal sources. Ambruster et al. (1983) proposed that some of the unidentified transients might be faint GRBs. As shown by Laros et al. (1984) about 2% of the GRB power is emitted in X-rays below 20 keV. This result suggested that GRBs with weak fluences $\sim 10^{-8}$ erg cm$^{-2}$ might appear as fast X-ray transients with fluences in excess of $10^{-10}$ erg cm$^{-2}$ in the range 0.5 to 20 keV. Searches in the data from the A-1 and A-2 Sky Survey experiments on HEAO-1 have shown that such transients occurred less frequently than $3 \times 10^{3}$ yr$^{-1}$ over the entire sky inconsistent with a $S^{-3/2}$ distribution (Ambruster & Wood 1986). Connors, Serlemitsos, & Swank (1986) gave a $2\sigma$ upper limit on the GRB rate of $4 \times 10^{4}$ yr$^{-1}$ above $10^{-9}$ erg cm$^{-2}$ (see their Figure 5). With the same idea Helfand & Vrtilek (1983) using data from the Einstein X-ray satellite searched for fast X-ray transients with fluences greater than $10^{-10}$ erg cm$^{-3}$. They concluded from their analysis that the upper limit to the rate of these events is $\sim 10^{5}$ yr$^{-1}$ at 1 keV (see their Figure 3). If $E_x/E_{\gamma} = 10^{-2}$ this point corresponds to $10^{-8}$ erg cm$^{-2}$ in the energy range of classical GRB detectors, still below the $S^{-3/2}$ curve. All these authors concluded that the frequency of fast X-ray transients gives some evidence in favor of a turnover in the log $N$–log $S$ curve. This was generally interpreted as another indication that cosmic GRBs had to be galactic.

Even though this short discussion seems to indicate that the interpretation of the log $N$–log $S$ curve is quite difficult, there were many attempts to compare burst observations with different galactic models—a disk or a spherical halo—sometimes with different conclusions (see, for instance, Fishman 1979, Jennings & White 1980). In the first works, GRBs were assumed to be standard candles, but it soon appeared interesting to test more realistic models with an intrinsic burst luminosity distribution (Cline & Desai 1976, Jennings 1982). Allowing an intrinsic luminosity distribution had a significant effect for halo and disk models. The impact was such that the halo and thick disk models, excluded for mono-luminosity bursts, became compatible with the observations. Nevertheless here again these results appeared in conflict with previous studies (Yoshimori 1978, Fishman 1979) which noticed the weak influence of the luminosity distribution on the log $N$–log $S$ curve.

With further studies, it rapidly appeared that the problem of the flattening of the curve at low fluence had to be questioned because this curve was affected by severe selection effects (Barat et al. 1982, Yamagami & Nishimura 1986, Higdon & Lingenfelter 1986, 1990, Mazets et al. 1980, Mazets & Golenetskii 1988, Jennings 1988). Among them, one selection effect which cannot be avoided is connected with the GRB detection method. In general for satellite observations, the telemetry is limited, preventing the recording of the gamma-ray flux continuously. The bursts are detected only when the counting rate of a detector increases above the background by 5–10
Figure 1.17. Burst fluence as a function of the burst duration indicating the selection effects for long burst with low fluence: GRBs with a low fluence and a long duration are not detected (Mazets & Golenetskii 1981b).

standard deviations, in an energy range $\Delta E$, within a time interval $\Delta t$, typically lasting 1 s or less. The counting rates used for the trigger are not directly related to the fluence, which is the total energy observed during the burst. For instance, for a given fluence $S$, if the burst is weak and long it will be below the threshold of the instrument. But, for the same fluence $S$ a shorter burst will trigger the instrument and the burst will be detected. Under these conditions for the lowest fluences accessible to one instrument, the short GRBs will be detected while the long ones will be missed with a high probability. This is illustrated in Figure 1.17 showing the measured fluence $S$ as a function of the burst duration using the data from the Konus experiment (Mazets & Golenetskii 1981b). This figure clearly shows the bias against observing long-duration, low-intensity bursts. This bias is particularly significant because GRBs have a quite large distribution of durations. This selection effect explains the flattening of the log $N$–log $S$ curve near the low $S$ threshold of the instruments (see in Figure 1.16 the results of IMP, Konus, and PVO). As an example, the demonstration of these selection effects has been done by Higdon and Lingenfelter (1986) on the Konus data. Taking them into account they arrived at the conclusion that Konus log $N$–log $S$ must be corrected and can be compatible with $S^{-3/2}$, in agreement with the isotropic distribution of GRBs. But these corrections did not remove the deficit of GRBs with very low $S$ values obtained with some balloon experiments.

Even if it was in principle possible to correct for the selection effects on the measure of the fluence, which were particularly significant for low $S$ values, other measures of GRB luminosity were proposed very early. Hence, it was suggested to use
$P_{\text{max}}$, the peak energy flux, and so to work with $\log N$–$\log P_{\text{max}}$ (Yamagami & Nishimura 1986, Higdon & Lingenfelter 1986, Mazets & Golenetskii 1981a, Mazets 1985). $P_{\text{max}}$ is more adapted since GRB detectors trigger when they detect an excess of photons above the background in fixed times and energy intervals. $P_{\text{max}}$ can be measured on a timescale and in an energy range matching those used for the trigger. Jennings (1985) used the Konus $\log N$–$\log P$, reaching the clear conclusion that Konus data treated to eliminate biases and artifacts, are statistically incompatible with sources that are nearby members of any known spherical or disk population. On the contrary they are consistent with earlier conclusions that bursts can be distributed in a large galactic halo.

While the peak energy flux is acceptable when only one instrument is considered, the difficulties come back when the results of different detectors have to be compared. This observation led to propose a new method, based on $C_p$ the peak counting rate in the burst and $C_{\text{lim}}$ the minimum counting rate needed to trigger the instrument. Figure 1.18 gives the cumulative size frequency distribution of Konus GRBs vs fluence, peak energy flux and peak counting rate. The introduction of $C_{\text{lim}}$ and not only $C_p$ as used by Paczyński and Long (1988) allowed taking into account background variations in orbit and so the variations of the threshold sensitivity.

Finally, after many years of controversy, Schmidt, Higdon, and Hueter (1988) proposed to use the $(V/V_{\text{max}})$ test, which measures for each GRB its position within the volume $V_{\text{max}}$ accessible to the instrument. For a given GRB, the quantity $V/V_{\text{max}}$ is easily connected with $C_p/C_{\text{lim}}$ defined in the previous paragraph. $C_p$, being the observed peak counts, depends on the source distance: $C_p = kr^{-2}$, for a Euclidean space. If $r_{\text{max}}$ is the distance at which a GRB source gives the minimum peak count $C_{\text{lim}}$ detectable by the instrument, we have $C_{\text{lim}} = kr_{\text{max}}^{-2}$.

$$C_p/C_{\text{lim}} = (r/r_{\text{max}})^{-2} \quad \text{and} \quad r/r_{\text{max}} = (C_p/C_{\text{lim}})^{-1/2}$$

![Figure 1.18.](image-url) Cumulative size frequency distribution of the KONUS bursts versus fluence, peak energy flux, and peak counting rate (Mazets 1985).
As the volume $V$ is proportional to $r^3$, we finally get: $V/V_{\text{max}} = (C_p/C_{\text{lim}})^{-3/2}$. For spatially homogeneous sources, $V/V_{\text{max}}$ is uniformly distributed between 0 and 1 with an average value of 0.5. The agreement on the use of $\langle V/V_{\text{max}} \rangle$ resulted, in 1990–1992, in the publication of the values measured by several GRB detectors (summarized in Vedrenne (1991), and Dezalay et al. (1994)). As expected, smaller detectors have $\langle V/V_{\text{max}} \rangle$ closer to 0.5, which characterizes a homogeneous distribution. PVO, with an effective area of 11 cm$^2$, had $\langle V/V_{\text{max}} \rangle = 0.46 \pm 0.02$ (Chuang et al. 1992, Hartmann et al. 1992). The Konus and SIGNE detectors, with effective areas of 50—60 cm$^2$, measured $\langle V/V_{\text{max}} \rangle = 0.45 \pm 0.03$ (Higdon & Schmidt 1990) and $\langle V/V_{\text{max}} \rangle = 0.435 \pm 0.02$ (Atteia et al. 1992) respectively. Phebus, with an effective area of 90 cm$^2$, measured $\langle V/V_{\text{max}} \rangle = 0.37 \pm 0.03$, while $\langle V/V_{\text{max}} \rangle = 0.4 \pm 0.08$ was measured for 13 bursts observed by HEAO-A4, with an effective area of 42 cm$^2$ in the energy range 80 keV–1 MeV (Schmidt, Higdon, and Hueter 1988). So for the largest detectors the value of $\langle V/V_{\text{max}} \rangle$ is significantly below 0.5, indicating that they have reached the boundary of the GRB population.

An interesting property discovered at that time was that detectors working at higher energies have lower $\langle V/V_{\text{max}} \rangle$ than detectors working at lower energies (Dezalay et al. 1994). This was explained later by the fact that intrinsically bright GRBs have harder spectra, allowing GRB detectors which operate above 100 keV to see them at larger distances. This again illustrates how the intrinsic properties of GRBs (luminosity, shape of the light-curve, spectral hardness) made the interpretation of their intensity distribution difficult.

Later V. Petrosian (1993) introduced a new way of analyzing the bivariate distribution of $C_p$ and $C_{\text{lim}}$ obtained directly from data after having demonstrated that $C_p$ and $C_{\text{lim}}$ are stochastically independent. This type of analysis which was performed for the SIGNE instruments, and for SMM and Ginga, showed that all these instruments measured a deviation from the slope $-3/2$, even before its latter confirmation with BATSE. For instance, for SIGNE experiments the deviation from the slope $-3/2$ was more pronounced and extended over a larger range than the one which would be deduced from the distribution of $C_p/C_{\text{lim}}$ given by Atteia et al. (1991) (see Figure 2 of Petrosian (1993)). All these results, combined with the log $N$–log $S$ tendency to flatten at low fluence and with the real isotropy of the angular distribution, cast some doubts on the galactic origin of GRBs.

Of course halo distributions were compatible with the uniform celestial distribution and the flattening of the curve log $N$–log $S$ (Nishimura & Yamagami 1985), but in order to have an isotropic source distribution, a large radius of the halo had to be considered: typically 100 kpc. The possibility of GRBs occupying a massive extended galactic halo was also considered by Jennings (1984). Such a distribution provided the best fit to the data but it did not correspond to previously identified galactic populations. Moreover, it demanded very large energy outputs ($\sim 10^{53}$ erg) for typical GRBs with a fluence of $10^{-5}$ erg cm$^{-2}$. Due to this enormous amount of energy, the possibility of having such a extended halo distribution was often rejected. For instance, it leads to large optical depth for photon–photon interactions, which are difficult to reconcile with the observed burst spectra extending to megaelectronvolts energies. Hence, Yamagami & Nishimura (1986) considered GRBs situated in the disk with a
scale height being typically 1 kpc, rather than in a halo; the log $N$–log $S$ flattening at low $S$ being explained by selection effects. They found that the log $N$–log $P$ curve (which is less sensitive to selection effects) obtained with Venera data is almost consistent with $P^{-3/2}$. So they concluded there is no conflict between the celestial distribution and the observed burst frequency curve in the disk model. This was also the conclusion of Higdon and Lingenfelter (1986) who considered that selection biases in flux and fluence could account for the reported deviations of the Konus size frequency distributions from that expected for a uniform spatial distribution of sources. However, the conclusion of Jennings (1988) was different: he found that after correcting for selection biases the curve log $N$–log $P$ still exhibited an obvious low $P$ turnover. Neither an isotropic homogeneous nor a nearby disk spatial distribution of monoluminosity GRB sources could provide statistically acceptable fits to this refined log $N$–log $P$. Introducing intrinsic burst luminosity distribution worsened the fits. So in the 1980s this problem of intensity distribution was puzzling and really an open issue.

Given all of this, it is very important to analyze now a different approach used by Hartmann, Woosley, and Epstein (1990) and Paczyński (1990). These authors carried out the very interesting exercise of calculating whether a galactic population of neutron stars (NS) could be associated with GRBs respecting the constraints we have just mentioned. They calculated the spatial distribution and kinematic properties of a sample of Population I NS born in a thin galactic disk with high velocities. They performed a Monte Carlo simulation of the birth of these NS and numerically integrated their subsequent orbits in a realistic galactic potential. The resulting burst source positions were used to calculate the statistics of the radial distribution, $\langle V/V_{\text{max}} \rangle$, the brightness distribution, log $N$–log $S$, the large-scale angular distribution on the sky (multipole moments) and the small-scale angular clustering properties. These statistics were then compared with the observed properties of GRBs to see whether the model assumptions were consistent with the observations. The conclusions of Hartmann, Woosley, and Epstein (1990) were the following: the angular distribution of GRB on the sky was consistent with that expected for Population I NS if the detectors sample events to a distance less than 150 to 2000 pc. Paczyński used a similar method (with a steeper galactic potential), but he obtained a velocity distribution of the NS with a larger low velocity component and the current sampling depth of the detectors was found to be 130 pc, much smaller than the value given by Hartmann, Woosley, and Epstein (1990). The distance limit he obtained at the 3σ confidence level was 600 pc (to be compared with the 2 kpc obtained by Hartmann, Woosley, and Epstein 1990). In both cases the sampling depth had to be 1 kpc at most if the GRBs were to be associated with galactic neutron stars. This was an important result because it implied that if GRBs were due to galactic NS, more sensitive detectors would undoubtedly reveal the galactic structure.

Overall, the results discussed in this section should have led us to severely question the galactic model but, as we have tried to demonstrate, the conclusions were never completely robust. At the same time, another way of establishing the GRB origin was considered: the search for GRB counterparts in the error boxes, the most
precise ones being obtained by triangulation, thanks to the Interplanetary Network. Let us have a look at some results of these searches.

1.6 GRB COUNTERPARTS

Most scientists were convinced that the mystery of the origin of GRBs would be solved with the identification of counterparts at longer wavelengths (optical, radio or X-rays). However, the search for counterparts requires accurate positions and before 1996 the only way to get precise (arcminute) positions was the triangulation method. This method uses the fact that GRBs are transient signals, and reconstructs their direction on the sky from the measurement of their crossing time at various points in the Solar System (Figure 1.13). The precision of localization is strongly dependent on the distances of the satellites (Figure 1.13) and on the precise knowledge of their position in space. Space probes going to Mars or Venus have been very useful in improving the precision of localization, sometimes down to about 1 arcmin. With the Franco-Soviet program SIGNE, and in collaboration with US satellites, a program of precise GRB localizations led by K. Hurley in France, and by T. Cline in the US was soon started; it was called the Interplanetary Network for GRB localization, or IPN. In the 1980s, several GRBs were precisely localized with this method, and significant efforts were devoted to searches for GRB counterparts. They concerned flaring (contemporaneous with the GRB), fading (which could remain visible in the days or weeks following the burst) or quiescent counterparts (always present).

All three types of counterparts were searched for in the optical, while only quiescent counterparts were searched for at radio and X-ray wavelengths.

Flaring counterparts. Two methods were used to search for optical emission simultaneous with GRBs. The first used ‘all sky surveys’ monitoring a significant fraction of the sky (>10%) for optical transients brighter than magnitude 5–10. Photographic plates and/or CCD images were recorded every clear night and examined a posteriori when a GRB was localized in the part of the sky which was monitored (e.g. Hudec et al. 1987). The requirement of spatial and temporal coincidence was mandatory in these searches, given the high background of optical transients due to meteors, planes, and satellites. The second method used archival plates taken in the direction of well-localized GRBs. The idea behind this technique was to look for past eruptions from GRB sources. The assumption that GRBs were emitted by galactic neutron stars led to the conclusion that they had to repeat to explain the observed GRB rate of a few hundred bursts per year, with only one galactic neutron star created every thirty years or so. This work required hundreds or thousands of archival plates of the GRB field and used plate collections available at the Harvard, Sonneberg or Ondrejov observatories (Schaefer 1981, Schaefer et al. 1984, Hudec et al. 1987, Hudec, Peresty, and Motch 1990, Atteia et al. 1985, Greiner et al 1987, Schaefer 1990). Many groups contributed to this work, which revealed some interesting possible counterparts. Nevertheless, after several years of deep studies the general feeling was that these searches had not provided convincing counterparts and that the few interesting possible cases could be explained by
background optical transients, possibly due to flare stars, meteors, planes, or satellite flashes. We now understand better all these negative results because all-sky surveys were not sensitive enough to detect the prompt optical emission from most GRBs (see Chapter 3), while archival plate searches were doomed to fail because GRB sources are not galactic and do not repeat.

**Fading counterparts.** Searches for fading counterparts were plagued by the long delays necessary to localize GRBs by triangulation, typically several weeks, due to the need to exchange magnetic tapes between researchers in the US, in France, and in the (former) Soviet Union. The identification of GRB afterglows in 1997 would be the result of the discovery of GRB fading counterparts, but in the 1980s these searches started too late. We know now that the fading optical emission from GRBs (the afterglow) decreases quickly and is only detectable above magnitude 21–22 during the few hours following the burst (up to 2–3 days after the GRB in the most favorable cases).

**Quiescent counterparts.** This is certainly the field which demanded the most effort and attention: quiescent counterparts to GRBs were searched at infrared (Schaefer et al. 1987), optical (Chevalier et al. 1981a, Motch et al. 1985, Hudec, Peresty, & Motch 1990, Seitzer, Schaefer, & Bradt 1983, Schaefer et al. 1984), X-ray (Grindlay et al. 1982, Pizzichini et al. 1986, Boër et al. 1988), and radio (Hjellming and Ewald 1981, Schaefer et al. 1989) wavelengths. These searches were based on deep observations of some small error boxes, but also on the statistical comparison of GRB positions with those of other astronomical objects (e.g. pulsars, quasars, X-ray binaries). These searches revealed a number of interesting objects in some of the error boxes. X-ray observations played an important role in these early studies. At the beginning of the 1980s several classes of galactic neutron stars were known, such as the radio pulsars, the X-ray pulsars and low-mass X-ray binaries (bursting or not). X-ray observations were very powerful in detecting accreting neutron stars in the Galaxy. In this context many observers were convinced that X-ray observations could be the missing link between gamma-ray bursts and their quiescent counterparts. The satellites Einstein and EXOSAT were used to observe several well localized GRBs but they did not find the expected X-ray emission (Pizzichini et al. 1986, Boër et al. 1988). For galactic sources, this lack of X-ray counterparts had important consequences, because it put strong constraints on the accretion rates, which had to be very low (of the order of $10^{-15} M_\odot \, yr^{-1}$), and on the neutron star temperature. Moreover, it excluded sources too close to the Sun (closer than a few hundred parsecs), whose thermal emission would have been detectable.

**GRB 781119.** To illustrate the efforts made to discover counterparts of well localized GRBs, we give the example of multi-wavelength deep searches in the error box of the strong GRB 781119. The Interplanetary Network with six satellites: Helios 2, Pioneer Venus Orbiter, Venera 11 and 12, Prognoz 7, ISEE C, and the Vela Satellites gave the opportunity to localize this burst with $\sim 5\, \text{arcmin}^2$ precision (Cline et al. 1981) and to undertake follow-on multi-wavelength searches. In the study of a collection of archival plates at the Harvard College Observatory, an optical transient (OT) with a duration shorter than 45 min was discovered inside the GRB error box by Schaefer (1981) on a photographic plate taken on November 17, 1928. VLA maps of
the field of the GRB obtained by Hjellming and Ewald (1981) showed three point-like
radiosources inside the error box but the authors did not claim definite evidence for
their association with the GRB source. Optical candidates were also searched in the
OT error box using the McGraw-Hill 1.3 m telescope (Schaefer & Ricker 1983). New
observations were made with the CTIO 4 m Telescope (Seitzer, Schaefer, & Bradt
1983) and the 1.5 m Danish Telescope at ESO (Pedersen et al. 1983). These authors
reported the detection of two very faint and variable sources within the error box of
the optical transient, but no proof could be given for their possible association with
the GRB. Finally, in the X-ray domain a quiescent source was detected with the
Einstein X-ray observatory at a location consistent with the 1928 optical transient
(Grindlay et al. 1982). But Boër et al. (1988) did not detect the source with
EXOSAT, this discrepancy might be due to the energy range of the EXOSAT
observation which was lower than Einstein’s. Later Boër et al. (1997) reported the
detection of two X-ray sources in the error box using an observation made with
the Rosat X-ray HRI experiment. One of them was the one discovered in 1981 by the
Einstein satellite, which had also been observed by ASCA (Hurley et al. 1996). The
other source was variable on the timescale of one year and was tentatively associated
with the GRB.

Figure 1.19a shows the error box of GRB 781119 with the possible counterparts
discussed above. Many studies of several other GRBs were unable to give a consistent
picture for any kind of counterpart to GRBs. Figure 1.19b provides an example
where no counterpart at any wavelength was reported for a well-localized GRB
(GBR 790613). This situation progressively led to a general consensus that GRBs
have no quiescent counterparts. But it should not be forgotten that all these
observations were done months or even years after the GRB explosion. Moreover,
the error boxes, even the smallest ones, were too large to allow indisputable
associations

In conclusion, when we look back at this long period of 15 years, we realize that
the efforts made, in particular by K. Hurley and T. Cline, to precisely localize GRBs
with the Interplanetary Network (IPN) were the starting point of a huge effort to
search for counterparts at many wavelengths. Astronomers like Schaefer, Hudec,
Pedersen, Greiner, or Motch were among the most active in this field, looking for
counterparts in the radio, optical, or soft X-ray domains, but in the end the results
were very frustrating. During this period, the IPN was the only way to obtain small
error boxes, and even if many researchers realized that counterpart searches were
made difficult by the long gap between the GRB explosion and its localization, delays
shorter than a few days were out of reach. In fact the majority of scientists involved in
the field believed that GRBs were galactic because it was more reasonable from an
energetic point of view. Some GRB characteristics such as cyclotron lines, very short
time variations, and lack of quiescent counterparts gave credit to the association of
GRBs with compact objects, mainly neutron stars. We will see that different models
have been proposed involving neutron stars in close binary systems, isolated neutron
stars accreting from the interstellar medium, neutron stars bombarded by comets, or
undergoing starquakes. Under these conditions GRB researchers were not always so
surprised to see that the searches for counterparts were unsuccessful.
Figure 1.19. (a) An example of a well-localized GRB, GRB 781119, thanks to many satellites and probes: Helios 2, Prognoz 7, Pioneer Venus Orbiter, Venera 11 and 12 (Cline et al. 1981). The figure shows two X-ray sources in the error box (Grindlay et al. 1982, Boër et al. 1997), and the position of an optical transient discovered by Schaefer in an archival plate recorded in 1928 (Schaefer 1981; see text). (b) Another possible example is the IPN localization of GRB 790613 with no object visible down to magnitude 21, the limit of the blue Palomar Observatory Sky Survey plates (Barat et al. 1984).
To conclude this first part on GRB observations, many properties of these events were established during the 20-year period following the discovery of these strange emissions. The temporal and spectral characteristics of GRBs were, however, unable to limit the large number of theories proposed to explain them (see, for instance, Hartmann 1991). Cyclotron lines observed by Konus and Ginga were nevertheless suggestive of high magnetic fields at the surface of neutron stars and therefore of a galactic origin for the bursts. However, strong constraints on galactic models were provided by the lack of quiescent X-ray counterparts, by the absence of repetitions and by the apparent spatial isotropy of the sources (Hartmann, Woosley, & Epstein 1990, Paczyński 1990). In the late 1980s, there was a strong need for more sensitive detectors which could detect a larger population of GRBs with a crude location of the faintest ones, a measure of both the spatial and the intensity distribution of a large population of GRBs which might definitely indicate their galactic origin. Finally the objective of searching counterparts as rapidly as possible was not the first priority even if, by the end of the 1980s, Ricker and colleagues were strongly defending this aim.

Before going on to the second step of the GRB saga with BATSE we will now say a few words on the different models involving mainly neutron stars, which were the most popular models at these early times.

1.7 The Neutron Star Paradigm

During the 15 years which preceded the launch of GRO much information on GRB characteristics was accumulated thanks to many satellites equipped with instruments specially designed for GRB studies, or by other missions like Vela, IMP-6 and 7, OGO 5, OSO 7, SMM, SAS 2, ISEE 3, HEAO-A. If we except the balloon experiments with large area detectors which were designed to study the weak bursts and to measure the log \( N \)—log \( S \) curve at low \( S \) values, the GRB detectors on satellites were small with typical areas from a few tens of square centimeters to about 100 cm\(^2\). Their sensitivity limits were of the order of \( S_{\text{min}} \sim 10^{-6} \) erg cm\(^{-2}\). They were omnidirectional and they had similar trigger algorithms. Many reviews devoted to GRBs have been written during this period. Among them, at the end of this period, a detailed analysis of GRB properties and a complete bibliography can be found in Higdon and Lingenfelter (1990).

In general GRB light-curves were observed to be quite variable and spiky with variations on very short timescales (usually less than 1 s, but sometimes as short as 10 ms). This pointed to sources with very small dimensions. Periods at the level of a few seconds were reported but they were never completely convincing. The GRB spectra could be fitted with various non-thermal models: OTTB (optically thin thermal bremsstrahlung), Inverse Compton, pair annihilation or synchrotron processes. But as the spectra are variable on very short timescales, very often the fits on timescales longer than a second could not really indicate the physical process at work during the burst. Nevertheless looking at the \( E \, dN/dE \) vs \( E \) spectra with a maximum
in the few hundred kiloelectronvolt range, it was clear that the maximum of the GRB power was emitted in the γ-ray domain and that this emission was always non-thermal. Important results concerning the presence of transient lines: cyclotron and redshifted annihilation lines, were first reported by the Leningrad group of Mazets. Even if these results were controversial, combined with the extremely short time variability, they represented strong arguments in favor of neutron stars as the sources of the GRB emission. With Ginga and its three examples of quite convincing cyclotron lines there were no doubts in the GRB scientific community at large that neutron stars were the sources of these violent, short, and energetic emissions. In this context, the lack of quiescent counterparts for well localized GRBs was significant (even the optical transients seen on archival plates were debated and suspected to be plate faults). This absence of counterparts could be explained if GRB sources were isolated neutron stars or neutron stars with very faint companions.

The neutron star hypothesis was also reinforced by the detection of an exceptional burst observed on March 5, 1979. This burst presented a short initial peak with a duration of 150 ms, a rise time of 0.25 ms indicating a compact source, a well-identified 8-s period (Mazets et al. 1979, Cline et al. 1980), and a spectrum softer than a typical GRB spectrum with a characteristic temperature $kT \sim 30$ keV. During the initial 4 s the spectrum showed an emission feature at $\sim 430$ keV, which could be interpreted as an annihilation line (at 511 keV) redshifted by the gravitational field at the surface of a neutron star. As this burst was very intense, it was observed by nine spacecraft and localized in a small error box in the direction of a young supernova remnant N49 in the Large Magellanic Cloud, at 55 kpc from the Sun (Evans et al. 1980, Cline et al. 1982). In the days, weeks, and years following the bright event of March 5, the same source emitted more bursts, with a total of 16 γ-ray bursts reported by Golenetskii, Ilinskii, and Mazets (1984). This source, which has been called GBS 0526-66 from its coordinates on the sky, seemed to be clearly associated with a neutron star. Even if the March 5 burst was early recognized as peculiar (Cline 1980), and very likely different from the usual population of GRBs or X-ray bursters (it will be later identified as a soft gamma-ray repeater), it was taken as another illustration of the diversity of the manifestations of neutron stars in the high-energy sky. Finally, this spectacular event strengthened the arguments in favor of a NS origin of GRBs.

All these observations led to considering neutron stars with strong magnetic fields as ideal candidates for GRB sources. The strong magnetic field was needed to explain the cyclotron lines and to confine near the surface of the star the very hot plasma which had to produce the GRB. Gravitationally redshifted annihilation lines could explain the features observed around 400 keV, even if there were numerous discussions about their reality. The absence of quiescent counterparts could mean that the neutron stars were solitary, as are most pulsars, or with faint companions not easily visible, for instance with red dwarf companions for a typical distance of 1 kpc.

In 1990, even if neutron stars appeared as the most suitable candidates, Paczyński and a few other researchers defended the idea that gamma-ray bursters could be at cosmological distances, like quasars, with redshifts around 1 or 2 (Paczyński 1986). The energy needed was comparable to the energy typically released
by a supernova in the electromagnetic domain, about 10^{51} \text{ erg}. This huge energy was also an argument to try to find solutions to the GRB mystery at much shorter distances, either in a quite small volume around us with a typical size less than a kiloparsec, or in an extended halo with a radius of \( \sim 200 \text{ kpc} \) (Shklovskii & Mitrofanov 1985, Atteia & Hurley 1986b, Jennings 1984), compatible with the isotropic distribution of the localized GRBs. Of course this last hypothesis assumed that a significant fraction of neutron stars with high velocity \( > 400 \text{ km s}^{-1} \) populated a large halo around our Galaxy. This might appear as an ad hoc solution to explain the isotropic distribution of GRBs and the values of \( \langle V/V_{\text{max}} \rangle \) obtained with different satellites. Of course the cosmological models were also well adapted because they explained the isotropy of the distribution, and predicted \( \langle V/V_{\text{max}} \rangle < 0.5 \) due to the expansion of the universe (which causes distant GRBs to appear fainter and softer than in a Euclidean universe). Nearby extragalactic models were excluded because there was a complete lack of correlations of the GRBs with the LMC or M31 (Atteia & Hurley 1986b). This discussion shows that, by the end of the 1980s, the cosmological models could have been in a much better position in spite of the huge energy they involve, if it were not for the absorption features in GRB spectra which were interpreted as cyclotron absorption lines on highly magnetized neutron stars.

So let us move on to the neutron star paradigm. As no indication of a definite counterpart existed despite important efforts made in multi-wavelength searches, there was still no way to choose between solitary neutron stars, such as old pulsars or close binary systems composed of a compact object and a faint companion. In a binary system, the source of the energy could be the accretion from a faint companion, while it could be the accretion of interstellar gas or cometary material for isolated neutron stars. But the energy might also be connected with internal or surface rearrangements of the neutron star, which were commonly observed in young pulsars, like the Vela pulsar. In fact in the 1980s the most accepted models could be divided into three classes: phase transition or starquakes in a neutron star, sudden accretion of matter on a neutron star and thermonuclear runaway in connection with slow accretion of matter from a companion or from the interstellar medium. As a conclusion to this chapter, we briefly discuss these three classes of models. Comprehensive and complete reviews of GRB theories and radiation processes at that time can be found in Lamb (1984a,b, 1988), Liang (1989), and Hartmann (1991).

\textit{Starquakes}

Initially proposed by Pacini and Ruderman (1974a), and Bisnovatyi-Kogan et al. (1975), starquake models were later refined by R. Epstein (1988) and Blaes et al. (1989). The development of starquake models has greatly benefited from observations of energetic glitches in the Vela radio pulsar. They allowed a better understanding of the coupling between the superfluid core and the crust with angular momentum transfer from the inner parts of the crust to the outer crust (Pines et al. 1980). These starquakes might liberate enough energy to explain a GRB due to the seismic excitation of magnetic field oscillations with a transfer of energy in the magnetosphere and the generation of strong electric fields. The electric fields parallel
to the magnetic field would accelerate electrons which would be at the origin of the GRB. Elastic energy released by a starquake might also be directly converted into γ-rays by means of a shock. The temperature in the shock might be high enough to produce γ-rays (Fabian, Icke, & Pringle 1976, Ellison & Kazanas 1983). But an open question was: if GRBs are associated with starquakes why were none observed during Vela glitches? Moreover, since the yearly GRB rate is greater than 100 per year and the birth rate of galactic neutron stars is <0.1 per year, each neutron star should produce $10^3$ GRBs at least over its lifetime. Hartmann (1991) showed that this number has to be greatly increased, to reach perhaps $10^5$–$10^6$ events per neutron star, if we consider that GRBs are only produced by young, local neutron stars. In the absence of observed repetitions, the GRB recurrence time remained an open issue. Nevertheless, even if the conservative limit of $10^3$ GRBs per neutron star was considered, it was not established whether neutron stars have enough energy to assure such a recurrence rate. The same comment can be made for the next possible mechanism.

**Sudden accretion on a neutron star**

In these models the accreted material comes from a comet, an asteroid or an accretion disk. This kind of model has been studied by Harwit and Salpeter (1973), Newman and Cox (1980), Colgate and Petscheck (1981), and van Buren (1981), but it suffered from the fact that the probability of comet impact might be low (Liang & Petrosian 1986), that the distribution of comet masses remained poorly known—even if new observations of the Halley comet by the Vega Spacecraft (Sagdeev et al. 1986) had allowed them to deduce that the number of comets could be larger than previously estimated (Mitrofanov & Sagdeev 1990).

**The thermonuclear model**

Even if the possibility of producing GRBs might exist through the two previous models there was a large number of open questions. In fact a model which has been studied more deeply, and not only to explain GRBs, is certainly the thermonuclear model. In this model the energy comes from nuclear burning of accreted matter (Woosley & Taam 1976, Woosley & Wallace 1982, Hameury et al. 1982). This matter can be accreted from a companion star or from the interstellar medium if the velocity of the neutron star is small and/or the density of the surrounding is high. The accretion is at a slow rate and the neutron star is highly magnetized, which means that the matter is accreted at the polar caps and confined by the strong magnetic field. Due to the high gravity, there is a gravitational settling which separates hydrogen and helium. A hydrogen layer is formed at the top of the mixture of He and metals. At the base of the hydrogen layer, with density about $10^7$ g cm$^{-2}$ and $T \sim 10^7$ K, the electrons are degenerate. Thermonuclear burning of the hydrogen heats this layer. The hydrogen burns mainly through non-$\beta$-limited CNO cycle and through electron capture by protons. Beyond a critical mass there is no equilibrium and an hydrogen flash will follow. The critical (minimum) accretion rate is $\sim 10^{-15} M_\odot$ km$^{-2}$ yr$^{-1}$. When the temperature reaches $\sim 10^8$ K the helium layer explodes, producing the
power needed for the GRB. The whole helium layer is burnt very quickly within $10^{-2}$ s. The energy released in this thermonuclear runaway is transported up to the atmosphere by Alfvén and magnetosonic waves driven by over-stable magneto convection. In the atmosphere these waves dissipate their energy by short-scale reconnections of the distorted magnetic field, which generate an electric field in the optically thin region. The electrons accelerated by this field produce gamma-rays by inverse Compton scattering with black-body photons of a few kiloelectronvolts. The resulting gamma-rays of a few hundred kiloelectronvolts are strongly beamed along the magnetic field lines, half of them escaping freely, while the other half move downward. The downward-moving photons undergo Compton scattering on cold electrons and create a subpopulation of hot isotropic electrons which emit synchrotron radiation which is approximately thermal with a temperature corresponding to the typical energy of excited electrons: $kT_e \sim mc^2$. The spectrum emitted by this illuminated photosphere is therefore the sum of a black body with a temperature of a few kiloelectronvolts, and a synchrotron radiation considered as thermal with a temperature $kT_e \sim mc^2$. The energy released is in the range $10^{37} - 10^{39}$ erg, implying distances between a few hundred parsecs and 1–2 kpc. The accretion rate being $10^{-15} M_\odot$ yr$^{-1}$, if matter is focused onto a surface of 1 km$^2$ it leads to a steady X-ray flux of $3 \times 10^{-14}$ erg cm$^{-2}$ s$^{-1}$. For a distance of 1 kpc, this is much lower than the limits deduced from Einstein observations (Pizzichini et al. 1982). This scenario was proposed by Hameury et al. in a series of papers (Hameury, Heyvaerts, & Bonazzola 1983, Hameury & Lasota 1985, 1989, Hameury et al. 1982, 1985). In the last paper of the series, they explained how the narrow cyclotron absorption lines could be produced and the low values of $L_\gamma / L_\nu$ due to the hot corona which filters the soft photons, upscattering them to high energies. Other detailed analyses of the thermonuclear model can be found in Woosley and Wallace (1982), and in Woosley (1984).

We have to note that the thermonuclear model had been previously used with success to explain type-I X-ray bursts or rapid X-ray transients. For these various phenomena the accretion rate and the magnetic field strength of the NS play a decisive role (Woosley 1982, Woosley & Wallace 1982). Similarly, the novae attributed to white dwarfs are also explained by thermonuclear runaways, after accumulation of a certain amount of matter. For all these reasons this model was considered as less ad hoc than the others we have mentioned. Moreover, due to the very low accretion rates which were needed, it seemed that it could work for isolated neutron stars accreting from the interstellar medium and for neutron stars in binary systems with faint companions. Figure 1.20 gives an artistic view summarizing how the thermonuclear model could produce GRBs (Atteia & Hurley 1986a, Hurley 1989a). Figure 1.21 is an illustration of a magnetized neutron star, a source of GRBs, with an indication of the different physical regions likely to be responsible for the $\gamma$-ray emission (Liang 1989).

1.8 CONCLUSION

After a 15-year period of GRB observations a general consensus was emerging: GRBs were certainly galactic and associated with strongly magnetized neutron stars.
The absence of counterparts and the March 5, 1979 exceptional burst with its possible association with a supernova (SN) remnant N49 strengthened this idea. Nevertheless the collective properties of GRBs cast some doubts on this scenario because absolutely no anisotropy could be detected in their spatial distribution, and even the faintest bursts seen by satellite experiments did not seem to be concentrated in the galactic plane. At the same time the log $N$–log $S$ curve, or $\langle V/V_{\text{max}} \rangle$, appeared to indicate an inhomogeneous distribution of faint bursts. These difficulties were circumvented by considering either a nearby population of neutron stars with a scale height smaller than 1 kpc, or an extended halo, with a radius larger than 100 kpc, constituted of high-velocity neutron stars ejected from the galactic plane. However, this halo model seemed an ad-hoc construction to satisfy the problem of isotropy.

Finally, as the inflexion of the log $N$–log $S$ curve appeared significantly at quite low $S$ values, where the localization of bursts was not possible by satellite experiments, much was awaited from BATSE, a next-generation instrument to be placed on NASA’s Gamma-Ray Observatory. The decade of the 1990s, which is the topic of the next chapter, will be crucial because the much larger sensitivity of BATSE will allow the detection and crude localization of much fainter bursts. The spatial distribution of faint GRBs was the result most awaited from BATSE.

Another crucial issue which was still open was the identification of GRB counterparts. Even if coincident observations by Vela spacecraft (Terrell et al. 1982) and P78-1 satellite (Laros et al. 1984) had discovered the emission of GRBs in soft X-rays, these observations were only used to define the properties of GRBs in the soft X-ray domain and not to try localizing them more quickly and accurately. In fact, the Vela spacecraft with its X-ray collimated detectors allowed the first and
Figure 1.21. Schematic diagram showing the physical regions likely to be responsible for the GRB spectra with its two components: the continuum and the possible line features (Liang 1989).
simultaneous determination of a moderately small error box for a GRB. But the late analysis of the X-ray emission did not lead to a counterpart search on a short timescale after the GRB. For many years after their discovery the search for GRB counterparts was not expected to remain so negative. Facing these difficulties in identifying an indisputable GRB counterpart, it was only at the end of the 1980s that G. Ricker and his colleagues in the US, in France, and in Japan proposed a quite new concept of mission for fast GRB localization. The HETE mission (High Energy Transient Experiment, Ricker et al. 1988) included the identification of GRBs in the classical way (NaI detectors), the localization of GRBs in soft X-rays (2–20 keV) with an accuracy better than 10 arcmin and at UV wavelengths with a UV camera which was able to localize transient events with an accuracy better than 10 arcsec. All the instruments had a large field of view, ~2.5 sr. What was quite new and very prospective for the field, was the capacity to obtain these localizations in near real time with small VHF receivers distributed on the ground (in such a way that the satellite was always in view of one of them), and to disseminate them quickly to observatories. The major aim of the mission was to track GRB counterparts as early as possible after the GRB, to see if they might be more easily identifiable if searches started a very short time after the explosion.

1.9 REFERENCES


2

The BATSE decade

BATSE (for Burst and Transient Source Experiment) on Compton Gamma-Ray Observatory (CGRO) was certainly the most ambitious experiment specially designed to study GRBs. BATSE consisted of eight pairs of NaI detectors. Each pair included a Large Area Detector (LAD) and a Spectroscopy Detector (SD). The area of a Large Area Detector was 2000 cm$^2$. With a sensitivity 5 to 10 times better than previous GRB detectors, BATSE detected an average of 1 GRB per day during the 9-year life of CGRO (1991–2000). These detectors were oriented in different directions to cover all the sky, and to localize GRBs crudely, with the techniques already used by Mazets with the Konus experiments. Due to the large dimensions of the detectors, the spectrum and the time profile of each burst were analyzed with unprecedented statistics. A quite complete review of the GRB situation and the possible models after 5 years of GRO-BATSE observations has been given by Fishman and Meegan (1995).

Here we would like to analyze the most significant results concerning the characteristics of the GRBs: morphology and spectra, and their global properties: spatial and intensity distributions. For each topic we will emphasize the confirmations of the properties outlined in the first chapter and the new results obtained with BATSE.

2.1 MORPHOLOGY AND SPECTRA

2.1.1 Morphology

The large diversity of GRB morphology was of course confirmed. Moreover as a large number of bursts were observed with the same instrument, a tentative classification was made (Fishman 1993). Among the different classes, one subclass was called FRED (Fast Rise Exponential Decay) with a rise time much shorter than the fall time. Their energy spectra almost always evolve from hard to soft. We have given an example of such a burst in the first chapter (Figure 1.2). However, these subclass
analyses have not contributed to any progress in the understanding of the GRB mystery. The fine temporal structure already observed was confirmed by BATSE which reported sub-millisecond structure (0.2 ms) for instance in a very short burst (8 ms GRB 910711, Figure 2.1; Bhat et al. 1992). Precursor activity was observed in some GRBs. A precursor is defined as a first episode of emission which has a lower peak intensity than that of the main emission and is separated from it by a background interval at least as long as the main emission. Kosht et al. (1995) have studied this behavior and found that about 3% of the GRBs observed with BATSE satisfied this criterion. They also found no evidence that the characteristics of the main episode emission are dependent upon the existence of the precursor emission and no evidence that the precursor emission and main episode emission are the results of different burst environments or production mechanisms (Koshut et al. 1995). No persistent, strictly periodic, structures were reported (Kouveliotou et al. 1992), confirming the previous observations. The only exception was the March 5, 1979 event, with its 8-s period, which is not a typical GRB but a Soft Gamma-Ray Repeater (SGR). Search for millisecond periodic pulsations in 20 bright bursts from 0.016 to 33.3 ms have been also unsuccessful (Deng & Schaefer 1997).

Another characteristic, which has been confirmed, is the fact that at higher energy, the overall burst durations are shorter and the spikes inside a burst are also shorter. This has been quantified by Fenimore et al. (1995). They used 45 GRBs belonging to the bright class (Norris et al. 1994) to study the variation with energy of the width of the temporal peaks. Using bright bursts made it possible to suppress the statistical uncertainties and other effects due to the

Figure 2.1. An example of a very short burst seen by BATSE illustrating how short a burst can be and its high variability at the millisecond timescale (Fishman et al. 1992).
distance (including possible time dilation effects from the expansion of the Universe, see Section 2.4). They showed that the narrowing of the peaks with energy follows quite well a power law with an index: \( \sim -0.4 \), so that for two pulses measured at energies \( E_1 \) and \( E_2 = 6E_1 \) the pulse at energy \( E_2 \) appears approximately two times narrower \( (6^{-0.4} = 2) \) than that at energy \( E_1 \) (see their Figure 3). This was the first quantitative relationship between temporal and spectral structures of GRBs, confirmed later by Norris et al. (1996b) for long bright GRBs.

Nemiroff et al. (1994b) have shown that long bursts (the short bursts were not considered in their analysis) are time-asymmetric which means that bursts and/or component structures rise in a shorter time than they decay on all timescales tested and for all energy bands tested. Barat et al. (1984) had already shown, for short single bursts seen by SIGNE instrument on board Venera that the rise to decay time ratio is \( \sim 0.32 \). Mitrofanov et al. (1996) and Norris et al. (1996b) have also confirmed this asymmetry of burst profiles.

Using 41 bright GRBs with durations longer than \( \sim 1.5 \) s Norris et al. (1996b) suggested what they called the pulse asymmetry/energy-shift paradigm (their Figure 15). Pulse shapes range from narrow and symmetric with negligible centroid shift with energy, to wide and asymmetric with centroid shift with energy. This pulse paradigm is a kind of unifying principle between short and long bursts. They also observed that the entire range of pulse widths extends from 10 ms to 2 s whereas the GRB durations extend over five orders of magnitude from \( \sim 10 \) ms to \( \sim 1000 \) s. For these bright GRBs another property is the distribution of intervals between pulses which exhibits a broad maximum near 1 s; such a characteristic timescale between pulses had been already noticed by Desai (1981). No evidence for photon bunching was reported in very bright GRBs (Mitrofanov 1989).

The excellent statistics of BATSE light-curves allowed so many studies of the GRB temporal properties that we cannot mention all of them. The large temporal variability of most GRBs was the object of many studies. Among them Lestrade et al. (1994), introduced a parameter which is a measure of the variability of a burst profile and allows one to quantify the ‘spikiness’ of bursts. They concluded that there is no indication of two separate populations on this sole basis (see Section 2.1.3). GRB light-curves were also analyzed from a statistical point of view. Stern and collaborators showed that the average autocorrelation function of GRBs is well represented by a stretched exponential, evidencing the self-similar properties of GRB light-curves. They proposed a simple ‘chain reaction’ mechanism which reproduces quite well the diversity of GRB light-curves with seven parameters (Stern 1996, Stern & Svensson 1996, Stern, Poutanen, & Svensson 1999, Beloborodov, Stern, & Svensson 2000).

Other authors reported that several temporal parameters (the rise and the decay times, the FWHM of the pulses and the separation between pulses) follow a log-normal distribution (McBreen et al. 2001, Quilligan et al. 2002), and showed that the cumulative light-curves of GRBs are typical of relaxation systems (McBreen et al. 2002).

The bimodality of GRB durations was well established by BATSE on hundreds of GRBs observed with the same instrument, and the same triggering characteristics (energy range and trigger time). This is illustrated in Figure 2.2, showing the duration distribution of 222 GRBs from the BATSE catalog. The duration $T_{90}$ corresponds, for a given burst, to the time over which 90% of the burst counts are recorded. $T_{90}$ is the duration of the interval starting when 5% of the counts have been recorded and ending when 95% of the counts have been recorded (Fishman et al. 1994). This is a way to have a measure of duration which is independent of the intensity for a given instrument. A clear bimodal distribution can be seen, two broad peaks appear centered around 0.3 s and 20 s with a minimum at about 2 s. The hardness of GRBs has also been calculated and represented as a function of their duration (Figure 2.3). The shorter bursts tend to have harder spectra than the long one (Kouveliotou et al. 1993). Such a difference between short and long bursts was also reported by Dezalay et al. (1992), and by Lestrade et al. (1993), using the data from Phebus experiment on the Granat mission.

Moreover, using Phebus data and a BATSE data set limited to events observable by Phebus, Dezalay et al. (1996) found a positive correlation between hardness and duration for long bursts ($T_{50} > 2$ s), while no such correlation was present for short GRBs. Since the two groups behave in different ways, this was an additional indication supporting the existence of two classes, and not a simple anticorrelation between duration and spectral hardness. On the other hand it appears that both short and long BATSE bursts have the same isotropic spatial distribution. A correlation between hardness and duration was also found by Horack and Hakkila (1997) among the 50 brightest bursts of the BATSE 3B Catalog (see their Figure 10).

BATSE bursts were also used to discuss the possible systematic effects affecting individual $T_{90}$ measurements (Koshtut et al. 1996). In addition, these authors compared the BATSE distribution with those obtained with other instruments: the SIGNE experiments and the Phebus instrument on the Granat satellite. They showed the coherence of these distributions when the triggering conditions are taken into account (timescales and energy range of the trigger).
2.1.2 GRB spectra seen by CGRO-BATSE

One characteristic of GRB spectra is their high-energy emission, which appears more clearly when $E^2 dN(E)$ is plotted vs $E$ (see Figure 2.4). This presentation, which shows the emitted power per energy decade, presents a maximum around a few hundred kiloelectronvolts. This maximum is a fundamental characteristic of GRBs, and it has been called $E_{\text{peak}}$ or $E_p$. Analyzing 54 GRBs, Band et al. (1993) found that the time-averaged spectra can be represented by the following empirical function:

at low energy, a power law with an exponential cutoff:

$$N(E) = AE^\alpha \exp - (E/E_0)$$

at high energy, a steeper power law:

$$N(E) = BE^\beta \quad \text{with } \alpha > \beta$$
Figure 2.4. Upper panel: An example of a GRB spectrum extending to ~10 MeV. GRB 910601 was observed simultaneously with BATSE, OSSE, Comptel, and Egret (Share et al. 1994). Lower panel: GRB 940217 observed by Egret, Ulysses and BATSE. Its remarkable spectrum extended into the GeV energy range as shown by Egret. The highest-energy point is based on a single 18 GeV photon. Points corresponding to the BATSE spectroscopy detector working between 30 and 2000 keV are shown with those obtained by the Egret total absorption shower counter (TASC) working between 1 and 200 MeV. The inset gives the deconvolved TASC and Spark chamber spectra (30 MeV–5 GeV) of the delayed emission. For more details see Hurley et al. (1994a).
The transition between the two functions occurs at $E_{\text{break}} = E_0(\alpha - \beta)$, while $E_p$ is defined by $E_p = E_0(2 + \alpha)$. Even if this general form fits the data from 20 keV to a few megaelectronvolts, there is not a universal average burst spectrum. In fact, a wide range of parameters is needed to fit all GRB spectra, with typical values centered around $\alpha = -1$, $\beta = -2.3$, and $E_0 = 150$ keV, close to the values discussed in Chapter 1. $E_0$ in particular is widely distributed from below 100 keV to about 1 MeV (see Figure 8 in Band et al. 1993), and the asymptotic slope of the low energy power law, $\alpha$, has been observed to vary with time rather than remaining fixed at its time integrated value. Band et al. (1993) reported a linear anticorrelation between $\alpha$ and the inverse of $E_0$ (see their Figure 3). Moreover, a correlation between $E_0$ and $E_p$ has been established in the time-resolved spectra of some GRBs (Crider et al. 1997) but this is true for the sample tested by the authors and it is not an universal property. This correlation is not incompatible with the anticorrelation of $\alpha$ with $E_0$ established by Band et al. (1993), because $\alpha$ is involved in the transformation between $E_0$ and $E_p$. Finally, no correlation was found between spectral parameters and burst morphology or location (Band et al. 1993). Pendleton et al. (1994b) reported no evidence for the presence of distinct burst classes based on their spectral hardness alone. So, representation of the continuum representation with four parameters appeared as a good way to characterize GRB spectra but it has no direct implications on the underlying physical processes.

Later Pendleton et al. (1996a) identified two different types of spectral emission which cannot be considered as two classes of GRBs. A subset of bursts exhibit a lack of fluence above 300 keV and these bursts have luminosities an order of magnitude lower than the other type of bursts which have a significant fluence above 300 keV. The subset of bursts without high-energy emission also has a homogeneous intensity distribution. These two classes of events may correspond to the extremes of an intrinsic hardness-intensity correlation. The authors showed that both types of emission are common in many bursts indicating that a single source object is capable of generating both types of bursts. The difference in the bursts would then be due to the more or less intense manifestation of a mechanism common to all GRBs.

Concerning the very high-energy emission of GRBs, Egret on the Compton Gamma-Ray Observatory (CGRO) has allowed observing the high-energy (>100 MeV) emission of some bursts detected by BATSE (Dingus et al. 1994). For instance, Egret detected 16 $\gamma$-rays above 30 MeV simultaneous in time and direction with the bright GRB 930131, while 0.04 were expected (Sommer et al. 1994). For some GRBs, COMPTEL also participated in the study of the spectrum beyond 10 MeV (Hanlon et al. 1994, Ryan et al. 1994). As an example Figure 2.4 shows the spectrum of GRB 910601 from $\sim 50$ keV to $\sim 100$ MeV, using three instruments of CGRO: OSSE, Comptel, and Egret.

A remarkable new result from the CGRO was the detection by Egret of delayed high-energy $\gamma$-ray emission from GRB 940217 (Hurley et al. 1994a). Very high-energy photons (200 MeV–10 GeV) were observed up to 5000 s (1.5 h) after the GRB, which lasted $\sim 180$ s at 100 keV. The time distribution of these photons is the following: 10 photons during the prompt $\gamma$-ray burst, followed by 18 photons during the next 5500 s, including a 18 GeV photon 4700 s after the trigger. Figure 2.4 shows this
burst observed by Egret, Ulysses and BATSE (Hurley et al. 1994a). Such very high-energy emissions from GRBs were reported by EGRET for at least five other bursts.

Coming back to the other results of BATSE, the significant spectral variability during GRBs has been confirmed (Pendleton et al. 1994b, Ford et al. 1994b). In the very intense burst GRB930131 Kouveliotou et al. (1994) demonstrated the variability of the spectral hardness on a 2-ms timescale. For most bursts hard-to-soft spectral evolution is observed. The peak energy is correlated with the intensity, it softens over the burst as a whole and also within individual intensity spikes (Ford et al. 1994b, 1995). These authors also found that bursts in which the bulk of the flux comes well after the trigger tend to show less peak energy variability and are not as hard as bursts in which the emission occurs promptly after the trigger. But as noticed by Kargatis et al. (1994), using 16 GRBs seen by the SIGNE experiments on Venera 13 and 14, no universal characteristic of spectral evolution could be determined: most of the time it is hard-to-soft, but it can also be sometimes soft-to-hard, or chaotic, even if many bursts (the majority) show a trend of increasing hardness with luminosity.

Another general property reported by Dezalay et al. (1992) for SIGNE experiments and by Kouveliotou et al. (1993) for BATSE was the fact that shorter bursts have harder spectra (see Figure 2.3). A remark can be made on this property. It could be imagined that short bursts are in fact the brightest peaks in the light-curves of long bursts whose long faint emission remains undetected because it is below the background. To see if this is true it is interesting to compare the spectra of short bursts with the spectra of the peaks of long-structured GRBs. This has been done by Koshut et al. (1994, 1995), who concluded that the spectra of the peaks of long duration GRBs are softer than the spectra of short bursts. Here again, all these characteristics are true in general, but there are exceptions.

Finally, concerning the continuum spectra, it is important to mention the existence of detectors like the GBD on Ginga or Watch on Granat which were sensitive to X-rays below 10 keV. These detectors confirmed the paucity of X-rays (X-rays usually contain few percent of the fluence of the burst), the presence of precursors in some GRBs (Murakami et al. 1991), and the presence of an extended emission at low energy (Yoshida et al. 1989, Murakami et al. 1991, Castro Tirado et al. 1994b). Nevertheless, Strohmayer et al. (1998), analyzing 22 GRBs detected with Ginga, found that some of them had a significant fraction of their fluence in low-energy X-rays (2–10 keV), much larger than the few percent usually observed in GRBs. They measure an average X/gamma fluence ratio of 24% with 4 events radiating more than 40% of their fluence in X-rays. This observation would be later confirmed with BeppoSAX and HETE, as discussed in Chapters 3 and 4.

The spectral features (emission and absorption lines) reported by Konus (for both lines) and by Ginga (for absorption lines) were not observed by BATSE (Palmer et al. 1994, Band et al. 1994, 1995, 1996, 1997, but see Briggs et al. 1998). Several other papers using BATSE observations arrived at the same conclusion: even if one detector indicated a feature it was not confirmed by another one (Ford et al. 1994a). This is the strength of using multiple, identical detectors with high sensitivity. Based on Ginga observations Fenimore et al. (1993b) estimated that BATSE should be seeing about 5.5 GRBs with lines per year, but BATSE did not observe these
GRBs. Later, with TGRS and the use of a germanium detector performing fine spectroscopy of GRBs, negative results would also be obtained (Kurczynski et al. 2000).

So, why did Ginga observe such strong lines while BATSE did not? As we have mentioned in Chapter 1 (Section 1.3.2), many criticisms had been raised concerning these lines. This discrepancy was important because the ‘cyclotron’ lines were one of the strongest arguments supporting highly magnetized galactic neutron stars as the progenitors of GRBs. So after the negative results of BATSE for such a large sample of GRBs observed with more than one detector, it was reasonable to forget lines as an argument to support the NS paradigm.

In conclusion, after BATSE and its large GRB zoo, the most important discrepancy with previous observations was certainly the absence of spectral lines for a large and homogeneous population of GRBs, and thus the absence of any clear signature of magnetized neutron stars. Before examining the global properties of GRBs with BATSE which will allow reducing the possible classes of GRB progenitors, let us have a look at some spectro-temporal correlations which might be useful for understanding the physics of the GRBs.

2.1.3 Correlations between spectral and temporal properties

A lot of work has been done on this subject, and part of it has been mentioned in the previous sections. To illustrate the strong activity in this domain and to complete the previous analyses we will choose a few examples.

**Hardness–brightness correlation**

The spectral hardness (measured by \( E_p \) or by a hardness ratio) has been correlated with the apparent brightness of the BATSE bursts (measured by the peak flux or peak counts). It was found that low-intensity bursts (dim bursts) exhibited spectra with a lower \( E_p \) on average (Nemiroff 1994a, Mallozzi et al. 1995, Dezalay et al. 1997). The consequences of this hardness–brightness correlation were discussed by Dezalay et al. (1997). In particular this correlation implies that the intensity distribution and \( \langle V/V_{\text{max}} \rangle \) depend on the energy range of the instruments. Since GRB sources might be at cosmological distances the spectra of dim (distant) bursts should be redshifted to lower energies relative to those of bright (nearby) bursts. This could explain this correlation (see Section 2.4).

**Spectro-temporal evolution**

The evolution of \( E_p \) with the GRB fluence was analyzed by Crider et al. (1999). A linear decay of \( E_p \) vs energy fluence was reported for 41 pulses within 26 GRBs. Such relations have to be considered with caution since the evolution of \( E_p \) in bursts is often complex (Ford et al. 1995). Liang and Kargatis (1996) found that \( E_p \) decays exponentially with the photon fluence (their Figure 3). Fenimore (1997) has studied the spectro-temporal evolution of an ‘average GRB’ constructed by adding together 32 bright BATSE GRBs, aligned at their peak. He finds that the peak energy is linearly proportional to the intensity.
Let us go now to the global properties of GRBs which were much awaited to illuminate the burning question: where are GRB sources situated?

### 2.2 SPATIAL AND INTENSITY DISTRIBUTIONS

#### 2.2.1 Angular distribution

Figure 2.5 gives the distribution in galactic coordinates of 1005 GRBs from the 3rd BATSE catalog (Briggs et al. 1996). While the isotropy of the distribution had been reported previously (e.g. Atteia et al. 1987), BATSE brought two new crucial features: the sensitivity of the instrument and the large size of the sample. With BATSE, one could check the isotropy of the weakest GRBs, too faint to be localized or even detected by previous instruments (Meegan et al. 1992, Fishman et al. 1994, Briggs et al. 1996, Paciesas et al. 1999).

The calculations of the dipole and quadrupole moments of BATSE GRBs were done by Briggs et al. (1996), taking into account all known observing biases. Again, and with a much larger confidence than before, there was no evidence of any anisotropy. The dipole moment $\langle \cos \theta \rangle$ differs from the value predicted for isotropy by 0.9 standard deviation and the observed quadrupole moment $\langle \sin^2 \theta - 1/3 \rangle$ by 0.3 standard deviation. In fact, GRBs are much more isotropic than any observed galactic population. This result is, of course, strongly in favor of cosmological distances, even if other distance scales can be accommodated as we will see below. The large number of GRBs made it possible to quantify this isotropy for different subsets of GRBs, again with no evidence for anisotropy (Meegan et al. 1992). The conclusion was definite, the celestial distribution of the weakest GRBs is perfectly isotropic, just like the celestial distribution of bright GRBs. This is true also for the short GRBs which are just as isotropic as long GRBs.

![Figure 2.5](image-url)  
**Figure 2.5.** Sky distribution of the 1005 GRBs in an Aitoff–Hammer projection in galactic coordinates regardless of the trigger energy range (Briggs et al. 1996). The apparent isotropy of the distribution has been confirmed by the calculation of its first moments.
For instruments like BATSE, with a localization uncertainty of several degrees, repeating sources may appear as clusters of GRBs, and the angular correlation function of the sources may be used to constrain the repetition rate. The detection of repetitions was very important because, if it was established, it could constrain the models. In particular, it would provide strong support for galactic models which are less demanding in energy, while cosmological origin of GRBs would involve catastrophic events which do not repeat. Quashnock and Lamb (1993b) and Wang and Lingenfelter (1995) found some evidence for repetition. These last authors indicated the presence of an excessive number of pairs of GRBs clustered in both space and time, and suggested that these bursts arise from repeating sources. Petroian and Efron (1995) found marginal evidence for bursts close on the sky occurring within 4–5 days of each other. Analyzing the angular distribution of 585 bursts, Meegan et al. (1995) found that the data are consistent with the hypothesis that γ-ray burst sources do not repeat (see also Brainerd et al. 1995, Narayan & Piran 1993, Rutledge & Lewin 1993). Schaefer (1994a) gave a critical analysis which made it possible to understand the discrepancies reported in the searches for possible repeaters. This point would be settled when the cosmological GRB origin was established (see Chapter 3).

### 2.2.2 Intensity distribution

We have seen that if the total fluence $S$ is used to obtain the intensity distribution, many biases and selection effects have to be corrected. So it was rapidly preferred to use peak energy fluxes (in erg cm$^{-2}$ s$^{-1}$) or peak photon fluxes (in ph cm$^{-2}$ s$^{-1}$). The maximum count rate $C_{\text{max}}$, and $C_{\text{max}}/C_{\text{min}}$ (see Chapter 1, Section 1.5) have also been used. But it has been noted that even if these quantities minimized the corrections it was difficult to connect them with physical units and to develop physical models of spatial distribution. To avoid these difficulties Pendleton et al. (1996b) preferred to use the peak flux distribution. Figure 2.6 indicates the peak flux distribution for 796 GRBs seen by BATSE. It is clear that for values of the peak flux $P < 5$ ph cm$^{-2}$ s$^{-1}$ there is a deviation from the $P^{-3/2}$ law shown as a dashed line (remember that this law applies to sources distributed homogeneously in a Euclidean space). As the number of localized bursts increased, the flattening below $\sim 5$ ph cm$^{-2}$ s$^{-1}$ (about $10^{-6}$ erg cm$^{-3}$ s$^{-1}$) was definitively confirmed with a slope close to $-0.6$ at the faint end of the distribution (Paciesas et al. 1999, and their Figure 6). A beautiful work was also done by Fenimore et al. (1993a), who combined BATSE results with those of PVO obtained after nearly 10 years of operation. Of course due to the small area of the detector, PVO detected the most intense bursts which are the most rare and which were not so numerous in the BATSE data one year after its launch. Their Figure 7 shows the log $N$–log $P$ distribution from combined BATSE and PVO data (Fenimore et al. 1993a). There is a very good agreement between the two instruments in the overlapping region, the curve at high flux is measured with greater precision and again the flattening at low $P$ values is visible without any doubt.

If we recall the discussions reported in Chapter 1 on the flattening of the log $N$–log $S$ or log $N$–log $P$ curves at low $S$ or low $P$, the connection between balloon results (at low fluence) and satellite observations (at high fluence) was not trivial. This
Figure 2.6. Integral log \( N \)– log \( P \) distributions for the standard 50–300 keV trigger energy range for peak flux measured on a timescale of 1024 ms. The dot-dashed line near threshold represents the correction for instrumental trigger efficiency (Pendleton et al. 1996b). The departure from the \( P^{-3/2} \) curve is clearly visible, well above the level at which threshold effects start to be important (around 0.3 ph cm\(^{-2}\) s\(^{-1}\)). So, the inconsistency with the \( -3/2 \) power law was definitely established with BATSE, proving the spatial inhomogeneity of the GRB population.

explains why the flattening of the curve below \( S^{-3/2} \) (or \( P^{-312} \)), even if it was observed before BATSE, was not considered as completely demonstrated. Moreover the flattening was observed in a region where few bursts were localized and it was difficult to get their angular distribution reliably. Of course this is not the case with BATSE and this is certainly the most fundamental result of the mission: the angular distribution of weak bursts which are in the flat part of the log \( N \)– log \( P \) curve is also isotropic. This means that the volume observed by BATSE reaches regions where there is a deficit of bursts and yet the burst distribution is isotropic. So without further discussion it is clear that a galactic disk distribution of GRBs sources has to be definitively rejected. However, before the examination of the consequences of this essential result let us complete the analysis of the intensity distribution with the \( \langle V/V_{\text{max}} \rangle \) test that we have explained in Section 1.5. Remember that if the distribution is homogeneous in a Euclidean space this ratio is distributed uniformly between 0 and 1, with an average value of 0.5. The first value of \( \langle V/V_{\text{max}} \rangle \) measured by BATSE was \( \langle V/V_{\text{max}} \rangle = 0.348 \pm 0.024 \), with only six months of data. Later, this number was completely confirmed with a much larger GRB number. With the BATSE 3B Catalog (more than 1122 bursts) one finds \( \langle V/V_{\text{max}} \rangle = 0.329 \pm 0.011 \), which is 15\( \sigma \) away from the value of 0.5 (Pendleton et al. 1996b). We have already given the values obtained with SMM, Ginga, Phebus and PVO respectively \( \langle V/V_{\text{max}} \rangle = 0.40 \pm 0.025 \), \( 0.35 \pm 0.035 \), \( 0.376 \pm 0.017 \), and \( 0.46 \pm 0.02 \) (for a complete set of \( \langle V/V_{\text{max}} \rangle \) values, see Hartmann and The 1993). It was clear, even before CGRO, that the \( \langle V/V_{\text{max}} \rangle \) test indicated a distribution which was not homogeneous in Euclidean space, because all the values were below 0.5, with the exception of PVO. But it was not surprising that
PVO, which was only sensitive to the strongest bursts, could not see the boundary of the local homogeneous distribution.

In the next paragraph we will discuss the consequences of these fundamental observations of BATSE on the possible spatial distributions of GRB sources. Before that, however, the reader may be curious to know how the intensity distribution of the short bursts (see Section 2.1.1) compares with the intensity distribution of long GRBs. While the two populations show the same level of isotropy, they have marginally different $\langle V/V_{\text{max}} \rangle$. Kouveliotou et al. (1993) found $\langle V/V_{\text{max}} \rangle = 0.367 \pm 0.03$ and $\langle V/V_{\text{max}} \rangle = 0.302 \pm 0.038$ for long and short bursts respectively. Mao, Narayan, and Piran (1994) showed that BATSE data could be accommodated by two populations of long and short bursts with the same spatial distribution. Interestingly, when they assumed that each population was made of mono-luminosity bursts, the two classes seemed to have the same peak luminosity to within a factor of 2. This appeared quite remarkable when we know that the duration of the short and long bursts differ by a factor of $\sim 50$ and that the energy of a typical short burst is smaller than that of a typical long one by a factor of $\sim 20$ (assuming identical spatial distributions of the two populations). We will see, however, in Chapter 3 that GRBs have a very broad intrinsic luminosity function, invalidating these conclusions.

### 2.3 Possible Spatial Distribution of GRB Sources

Before presenting a short analysis of the possible GRB models allowed by BATSE, it is interesting to refer to the paper of Nemiroff (1994), which provides a list of more than 100 GRB models (his Table 1) with a discussion on the possible origin of GRBs. A lot of these models involved a galactic disk population of neutron stars, so after BATSE many of them had to disappear from the list. To summarize this exhaustive analysis we can say that before BATSE the favorite candidates were highly magnetized neutron stars situated in the galactic disk. With BATSE, the situation was more complicated with three large classes of models which were debated and at least two of them which could not be definitively rejected.

#### 2.3.1 Single component galactic models

For these models highly magnetized neutron stars are still the favorites to explain GRBs but the question is to know in what volume they are confined. They might be situated:

(a) *In the galactic disk*. As we have seen, in this case there is a conflict between the isotropic distribution of localized GRBs and the deficit of weak bursts observed with the number-intensity curves and the $\langle V/V_{\text{max}} \rangle$ test. Hakkila et al. (1994b) have shown that disk models cannot be saved by allowing bursts to have broad luminosity functions. Such models cannot eliminate the anisotropy that must accompany the inhomogeneity of galactic disk distributions. Their conclusion was clear: no disk distribution can be made compatible with BATSE results. Mao and Paczyński
(1992) have also shown that even the early BATSE determination of the galactic quadrupole moment and \( V/V_{\text{max}} \) were inconsistent with any disk geometry. Clearly the much higher sensitivity of BATSE had allowed excluding these classes of models. Since no single population of sources confined to the galactic disk, the galactic halo or localized in our spiral arm could explain the BATSE observations, Hakki et al. (1994a), among others, proposed an extended spherical galactic corona. This was the other way to try to preserve the NS paradigm, which we discuss now.

(b) In an extended galactic halo or corona. This structure could be populated with galactic NS ejected with large velocity (>600 km s\(^{-1}\)) from the disk (Harrison, Lyne, & Anderson 1993). This scenario was supported by the discovery of high-velocity radio pulsars well above the galactic plane. These pulsars, with velocities which can reach \( \sim 1000 \) km s\(^{-1}\), are not gravitationally bound to the galactic disk. In this scenario, it is, however, necessary to explain why the most numerous population of NS situated in the disk do not give GRBs. Li and Dermer (1992) supposed a delayed turn on time for these NS of typically 30 Myr. Another possibility was to consider NS which were directly born in the halo (Hartmann 1992, Brainerd 1992b, Eichler & Silk 1992). Fabian and Podsiadlowski (1993), for instance, proposed that GRB sources could be ejected from the Magellanic Clouds, producing an extended halo around the Galaxy, nearly homogeneous out to 50 kpc.

Extended halo models have also been proposed by Li and Dermer (1992) and Shlovskii and Mitrofanov (1985), and well studied by Hartmann et al. (1994). Based on the statistical properties of GRBs a galactic neutron star origin could not be abandoned. But the models suggest that only a small fraction of galactic NS resides in the halo. On the other hand, the angular isotropy of GRBs indicates a strong domination of halo sources, implying enhanced bursting activity for halo stars and/or suppressed activity for disk stars in order to explain why the most numerous population of NSs situated in the disk did not give the majority of GRBs. The injection of high-velocity NS from the disk as proposed by Shlovskii & Mitrofanov (1985) could also generate acceptable distributions but required selection effects which could reduce the burst detectability while these stars were still close to the galactic plane\(^1\) (Hartmann et al. 1994) or a delayed turn-on time for these NS of typically 30 Myr (Li & Dermer 1992). Overall, the direct injection from parent bodies already situated in the galactic halo alleviated some of the problems associated with disk injection (Hartmann 1992) and led to a much more isotropic distribution.

Another important question was the size of the halo needed to fit the isotropy of the observed bursts. Mao and Paczyński (1992), using a large variety of luminosity functions for the sources and different values of the halo core radius, concluded that the BATSE results were compatible with a halo distribution if the halo core radius was very large \( R_c > 18 \) kpc (3\( \sigma \) limit). This core radius was introduced to determine

\(^1\) It was suggested for instance that high-velocity NS could emit GRBs beamed along the axis of their motion (Duncan & Li 1997). Such a scenario would lead to GRBs which would become detectable only when the sources, born in the galactic disk, are far from their birthplace.
the density distribution of Galactic halo objects: \( n(R) = n_0/[1 + (R/R_c)^\alpha] \), \( R_c \) is the core radius into which the density is constant, and \( R \) is the distance from the Galactic Center. The values \( \alpha = 3 \) or \( \alpha = 2 \) correspond respectively to luminous and dark matter distributions. The density of luminous objects like stars falls off as \( R^{-3} \), and the density of dark matter falls off as \( R^{-2} \) (Paczyński 1991b). Brainerd (1992b) found that the core radius could be much smaller \( (R_c \sim 5 \text{ kpc}) \) if GRBs were not standard candles.

Such distributions are not observed for any galactic class of objects. So, a new population is required. Neutron stars or other objects ejected from the galactic disk would not produce a uniform number density over the inner \( \sim 18 \text{ kpc} \), no matter what the ejection velocity, since the disk is strongly concentrated toward the Galactic Center. In spite of these objections this kind of distribution was intensively studied. Hence Hakkila et al. (1993) using 400 BATSE GRBs found that the core radius had to be larger than 34 kpc with an extension of the halo to at least 140 kpc. In fact it was difficult to pin down these two parameters, the core radius and the sampling distance, without connecting them to the value of \( \alpha \) which appears in the radial dependence of the source density (defined in the previous paragraph). Hakkila et al. (1993) gave the constraints on the core radius as a function of the sampling distance for different values of \( \alpha \). Their Figures 1 and 2 give the acceptable parameter space for two cases assuming mono-luminous GRBs: a corona/extended halo model and a two population model with a corona and sources in the local disk or halo. For instance for \( \alpha \) between 2 and 3, they found that the core radius is approximately between 50 and 100 kpc and the sampling distance between 150 and 400 kpc (their Figure 1). Later, Hakkila et al. (1995) returned to extended galactic halo models; they found that the core radius has to be situated between 50 and 80 kpc, so that the Galactic Center and/or M31 (respectively situated at 8.5 and 780 kpc) do not contribute significantly to the total number of observed bursts. Moreover, they found that bursts originating in extended galactic halos must have very narrow luminosity ranges with \( L_{\text{max}}/L_{\text{min}} \) less than 5. As noticed by the authors it is surprising that in spite of the large diversity in time and spectral properties of GRBs they need to have essentially the same peak luminosity.

Concerning the size of the halo Hartmann et al. (1993, 1994) concluded that a typical source distance of the order of 100 kpc is needed. This means a source luminosity of the order of \( 10^{42} \text{ erg s}^{-1} \), already very large with respect to galactic sources at typical distances of 1 kpc. Moreover, an upper limit to this distance had to be about 400 kpc due to the non-detection of an excess of bursts from M31. If \( \sim 10^5 \) objects are present in this large halo (assuming a ratio of about \( 10^{-3} \) between NS formed in the halo with respect to those formed in the disk), the total burst rate inferred from BATSE data (\( \sim 800 \) GRBs/year, Meegan et al. 1992) implied a typical recurrence time of the sources of the order of 100 yr, consistent with current constraints giving a recurrence time \( > 10 \text{ yr} \) (Hartmann et al. 1994).

In conclusion, even if these halo models could match the observed isotropy and give acceptable recurrence times, the sources did not correspond to any previously known class of galactic objects. Dark matter halos do not work and high-velocity neutron stars must have particular properties with respect to neutron stars in the disk.
of the Galaxy. Finally, the source of energy needed to produce GRBs from these halo NS remained to be found (e.g. fossil disks or very strong magnetic fields). Harding and Leventhal (1992), for instance, have shown that slow accretion of gas on high-velocity NS in the halo cannot be the origin of GRBs. Like Mao and Paczyński (1992), we can say that there are many disadvantages to and no positive evidence in favor of extended halo distributions. This is not far from the conclusion of Hakkila et al. (1995) who indicated that these new galactic models are perhaps as difficult to construct as cosmological ones. But before jumping to the cosmological models let us say a few words on the two-component models.

2.3.2 Two-component models

During this GRO period, even if the cosmological model was gaining ground, two-component models were also defended (Lingenfelter & Higdon 1992, Smith & Lamb 1993, Katz 1994, Higdon & Lingenfelter 1994). One of the goals of these models was to preserve a population of galactic NS which could explain the cyclotron lines seen by Ginga, even if no such lines were reported by BATSE for a large and homogeneous sample of spectra, casting some doubt on their reality.

The first model of Lingenfelter and Higdon (1992), with two galactic components arising from a single source population, was rejected at the 6σ level by BATSE data because of its quadrupole moment. In their revised model, Higdon and Lingenfelter (1994) claimed that the BATSE spatial and intensity distributions were consistent with halo distributions with core radii between 7.5 and 15 kpc, with typical source distances of 100 kpc, and a cutoff at 300 kpc, mixed with a local disk population at typical distances of 300 pc which can account for as much as ~25% of the observed BATSE bursts. This local population could account for the cyclotron lines which were observed by Ginga in a limited fraction of GRBs (less than 20%), but this is more difficult for the halo population with typical distances 100–1000 times larger. The luminosity would have to be 10^4 times the Eddington luminosity for a NS, unless the radiation is highly beamed. As noticed by Harding (1994b), static source models cannot apply since radiation pressure would drive a wind away from the neutron star. So, cyclotron line models seemed in danger since lines were produced in the static magnetized plasma atmosphere of a neutron star. Fortunately, Miller et al. (1991) have shown that these lines could also be produced in a relativistic, radiation-driven wind. Of course, if the lines can be produced in a relativistic, radiation-driven wind, even the neutron stars in the halo could produce them.

The compatibility of two population models with BATSE observations was disputed by Paczyński (1992b). He considered that introducing a disk population led to more severe constraints on the remaining bursts in the extended halo, and he concluded that two population models had to be rejected. Nevertheless, Hakkila et al. (1994a,b) found that a galactic disk distribution could be accepted if it does not contain more than 20% of the sources. Again, this means that the primary population accounts for the general isotropy and inhomogeneity seen by BATSE (see also Smith & Lamb 1993).

Smith (1995) returned to disk + halo models, considering galactic halo sources
producing the long/smooth bursts (80%) and disk sources producing the short/variable bursts (20%). This corresponds to the separation of GRBs into two classes reported in Section 2.1.1. A combination of an exponential halo plus an exponential disk was used to show that BATSE observations were consistent with such models. Their consequences were discussed: the number of disk and halo sources, their birth rates, their total energy release (one order of magnitude higher for halo sources over their bursting life-time), the much higher luminosity of halo bursts, etc. But these models have unanswered questions like (Smith 1995): Why are the two classes of short and long bursts consistent with isotropy? Why are disk sources not producing halo-type bursts?

Taking into account the same arguments which seemed to link some GRBs to galactic neutron stars, Katz (1994) introduced two populations: one at cosmological distances and the other galactic.

A major argument which casts doubt on two-component models was noticed by Briggs et al. (1996). The ratio of the distances of the two populations is at least a factor of 100, implying average luminosities which differ by a factor of $10^4$ or more. It would be very surprising if sources with such large different luminosities did not reveal themselves to be observationally separable into two classes. The observed bimodality in GRB durations might be the best indication of two classes but the short and long bursts are both consistent with isotropy and inconsistent with homogeneity (Kouveliotou et al. 1993). It is therefore difficult to believe that one type of burst is in the disk and the other in the halo (Smith & Lamb 1993). As Mao, Narayan, and Piran (1994) observed, it would be too much of a coincidence for the two populations to have the same statistical properties. Another argument which was not in favor of a nearby galactic population was the fact that the distribution of 60 well localized GRBs published by Atteia et al. (1987), and obtained after many years of observation, showed no anisotropy. Galactic disk models predicted that some galactic anisotropy should be detectable among the brightest bursts. However, as Higdon and Lingenfelter (1994) assumed a disk distribution with a maximum distance (300 pc) smaller than the scale height of the disk, the isotropy of nearby GRB sources did not really contradict their model.

A detailed comparison of BATSE results with various galactic models was made by Briggs et al. (1996). They calculated the dipole and quadrupole moments for each model and their deviations with respect to BATSE (the values are given in their Table 8). It follows very detailed discussions on these different models; in particular, the geometric halo model, high-velocity neutron star models and two-component models are deeply analysed with their constraints and their weaknesses. The clear conclusion is that BATSE observations disprove the previous paradigm of a galactic disk origin with magnetized neutron stars as GRB sources. Generally, no known galactic component, including the dark matter halo, is sufficiently isotropic to match the observational constraints (Briggs et al. 1996).

Finally, a good way to avoid an extended halo or two population models, which appear somewhat ad hoc, is to consider cosmological models. Even if they imply the release of huge amounts of energy in the $\gamma$-ray domain, they are perhaps the most natural ones. At least, they naturally account for the statistical properties of GRBs.
2.3.3 Cosmological models

Proposed by Usov and Chibisov (1975), and Prilutskii and Usov (1975), as early as 1975, they have been defended with strong conviction, mainly by Paczyński (see for instance Paczyński 1986, 1991a, Mao & Paczyński 1992). In these models the typical distance of the sources is at redshift $z \sim 1$ (about 3 Gpc), which means a typical source luminosity of $10^{51}$ erg s$^{-1}$, and the isotropy and the deficit of faint GRBs are naturally explained by cosmological effects. Sources which uniformly fill the observable volume of the Universe are perfectly isotropic and their density decreases due to the expansion of the Universe which attenuates GRBs and redshifts their radiation outside the energy range of BATSE.

As the results of BATSE considerably weakened the galactic models, a lot of theoretical works flourished to find a way to produce the energy needed by cosmological models and to account for the non-thermal emission of the bursts with emission at and above 1 MeV. One of the major difficulties, which exists also for sources in an extended halo, is the high optical depth for pair production in extremely luminous sources like GRBs. With the enormous energy which has to be liberated a fireball would be produced. This relativistically expanding pair plasma was introduced for the first time by Cavallo and Rees (1978), and its characteristic spectrum was outlined by Cavallo and Hortsman (1981). We will return to these fireballs in Section 2.6 and in Chapter 5, with all the developments made by the work of Mészáros and Rees when the cosmological model was confirmed after the remarkable observations of BeppoSAX.

In spite of the increasing weight of cosmological models, no distance was firmly attributed to a GRB, the searches for counterparts (see Section 2.5) leaving no convincing identification at any wavelength. Under these conditions it was natural to search in the GRBs themselves for some properties which could be the signatures of their cosmological distances. The tests which received particular attention were the searches for time dilation and spectral softening of dim bursts, and the searches for lensed bursts (Paczyński 1991a, 1992a).

2.4 SEARCHING FOR COSMOLOGICAL SIGNATURES

2.4.1 Time dilation

If GRBs are at cosmological distances the more distant bursts have to be fainter and receding faster (due to the expansion of the Universe). So, the weakest and farthest bursts must show larger time dilations (by an amount proportional to $1 + z$) than the nearer and more intense ones, they must be stretched and on the average they must be longer. The individual structures within bursts and the time intervals between these pulse structures have also to be stretched. The spectra of the weakest bursts must also be redshifted and softer. Due to the wide diversity of the GRBs (time structure and duration) time dilation can only be tested in a statistical way.
Norris et al. (1994, 1995) analysed burst samples in three peak intensity ranges. In their first (second) paper they took 131 (159) bursts out of 590 (775) GRBs, excluding bursts shorter than 1.5 s. They observed that the dim bursts are stretched by a factor 2.25 relative to the bright ones. If GRBs are standard candles at cosmological distances, this stretching is consistent with the source of the dimmest bursts being at redshift $z \sim 1–2$. The time dilation was later revised downward, to approximately 1.75, more in line with the redshifts expected from the brightness distribution (Norris et al. 1995), but still higher than the redshift $z \sim 0.8$ found for dim BATSE bursts by combining PVO and BATSE number intensity relations (Fenimore et al. 1993). Unfortunately, this anticorrelation of brightness with duration could have other explanations. As a caution, Norris et al. (1994) noticed that if the burst mechanism is such that the total energy for a burst is approximately constant then the release of energy at a slower rate corresponding to a lower luminosity would produce longer bursts. Dim and long bursts could thus be due to the nature of the physical processes at work in the burst.

Other studies by Davis et al. (1994) led to the same conclusion: the width of pulse structures within strong GRBs appears systematically shorter than those of the weak bursts: weak bursts have durations that are on average two times longer than the bright ones. They noted that the best fits to the number-intensity relation indicate a redshift $z \sim 1$ corresponding to the dimmest bursts used in their analysis.

Horack, Mallozzi, and Koshut (1996) proposed a method of assessing the consistency of time dilation measurements with both the brightness distribution of BATSE and various cosmological models. They used it to analyze the time dilation results of Norris et al. (1995). They noticed that a measured time dilation factor of 2.25 from two sets of bursts with peak fluxes of 16.2 and 0.78 photons cm$^{-2}$ s$^{-1}$ respectively corresponds to redshifts $z_{16.2} = 0.62$ and $z_{0.78} = 2.8$ for an Einstein–de Sitter cosmology. The bright bursts ($P = 16.2$) are obtained from the region of the distribution which is consistent with a $3/2$ logarithmic slope for the log $N$–log $P$ distribution, indicative of a homogeneous distribution of sources in a Euclidean (nearby) space. Assuming a non-evolving burst population, a redshift $z = 0.62$ for the bursts is incompatible with the brightness distribution, as this would produce a significantly different logarithmic slope than the observed $–3/2$ value (Horack, Mallozzi, and Koshut 1996). Even with the revised value of 1.75, these authors raised the possibility of an inconsistency with the brightness distribution of non-evolving bursts. This inconsistency was analyzed and it was noted that the consideration of energy-width corrections (as found by Fenimore & Bloom (1995), this is an important correction) exacerbates the difficulty, since bursts are found to be at higher redshift when the energy correction is applied.

This inconsistency was not the only problem, and even the time dilation measurements were disputed. Mitrofanov et al. (1996) for instance did not agree with Norris’ conclusions. Norris et al. (1994, 1995) made their quantitative analysis of the time profiles through the use of two tests: Haar wavelet transforms and the alignment of highest peaks technique proposed by Mitrofanov et al. (1994). To measure a possible dilation effect Mitrofanov et al. (1996) used their method of averaging bright and dim GRBs with the alignment of their principal peak. This method is well explained in
their paper: they define an average curve of emissivity (ACE) for a selected set of GRBs (the bright and the dim ones). For each time bin $t_i$ before and after the time of the highest peak they compute the normalized fluxes as $f_i = C_i/C_{\text{max}}$ (where $C_{\text{max}}$ is the counting rate of the highest peak of the burst) which are then averaged over the selected set of GRBs. The average dimensionless value is $F_i = \sum f_i / N$ where $N$ is the number of bursts which are considered in the averaging process. The values $F_i$ represent the average curve of emissivity (ACE). All the contributing sources are aligned at their primary peak. In the same way, the average spectrum can be calculated by taking into account a contribution from each burst which is proportional to its relative flux in each time bin. In the same way, one can construct the average hardness ratio of the GRB sample, this is the ACHR (averaged curve of HR). Figure 1 of Mitrofanov et al. (1996) illustrates the ACE and ACHR for the set of BATSE GRBs in the energy range 50–300 keV. The hardness ratio is defined as the ratio of counts in the energy ranges [100–300]/[50–100] keV. Incidentally, this technique is useful to determine average properties of a set of GRBs. For instance, this method made it possible to show that the rising edge of the average profiles is shorter than the falling one (2.3 s vs 4.2 s for a set of 338 GRBs). This asymmetry has been also observed by Nemiroff et al. (1994b). The ACHR (their Figure 1b) shows a hardness ratio higher before $t_0 = 0$ (the time of the highest peak) than after. Moreover a comparison of bright and dim subsets shows a higher HR for bright bursts than for dim ones during the entire burst duration. Bright events are harder than dim ones at all times. These properties have already been reported. Coming back to the stretching question, the results obtained with the ACE method by Mitrofanov et al. (1996) did not show a statistically significant difference between the average emissivity curves of bright and dim sets (their Figure 3 can be compared with the Figure 3 of Norris et al. 1994). The authors concluded that there is no real difference between the ACE of the bright and dim data sets. Koshut et al. (1996), looking for evidence of time dilation found that BATSE data were consistent with a redshift of 0, although the presence of time dilation could not be ruled out. The reason for the discrepancy between the analyses of Norris and Mitrofanov could be due to the different selection criteria used to define the two subsets. In particular Mitrofanov et al. (1996) introduced shorter bursts in their subsets. This explanation is illustrated by the results obtained by Mitrofanov et al. (1996), using the same sets of bright and dim bursts as Norris et al. (1994). The results are given in Figure 5 of their paper, and the stretching of dim bursts is apparent. This long discussion illustrates the difficulties connected with these analyses and the difficulties of reaching a conclusion.

Even if time dilation is present, it is not necessarily the signature of the cosmological distances of GRBs. For instance, Band (1994) made a critical analysis of the tests which were used by Norris et al. (1994) and he concluded that the time dilation observed by Norris et al. could easily be produced by a burst luminosity function. The duration and the peak photon luminosity might be inversely related if for instance the total photon emission is constant. The cosmological signature could also be diluted by a broad luminosity function. So, the time dilation can be cosmological but burst physics may also conspire to produce the same properties (Band 1994). Brainerd (1994) has shown that a lengthening of burst duration with decreasing flux is easily
produced by an inhomogeneous distribution of bursts in Euclidean space. Moreover
the flux at which the time dilation appears is also the flux below which the log N–
log P curve deviates from a $-3/2$ power law. So, he argued that the observed onset of
time dilation at fluxes below the break in the log N–log P is not evidence that burst
sources are cosmological (see also Wijers & Paczyński 1994). Finally, intrinsic burst
properties might dominate over the effects of cosmological expansion. For instance,
GRB models considering bulk relativistic motion may introduce a correlation of the
burst duration with its flux if the bulk motion has a privileged direction. This effect is
strong and may compete with cosmological redshift (Brainerd 1994).

Wijers and Paczyński (1994), analyzing samples of bright and faint GRBs from
the Second BATSE Catalog concluded that a duration change by a factor of 2 is
observed, confirming the results of Norris et al. (1994). They also found that the
duration difference between faint and bright bursts is explained by an overall shift
of the distribution without a change in shape, as expected for cosmological time
dilation. Concerning time dilation, Fenimore and Bloom (1995) observed that the
narrowing of GRBs at high energies significantly reduces the observed time duration.
According to Fenimore et al. (1995) GRBs observed at higher energies are shorter,
with narrower peaks. GRBs at high redshifts are observed at high energies in their
restframe and they are thus intrinsically shorter; this effect partly compensates for the
time dilation, which is thus less strong than the $(1 + z)$ factor naively expected. When
it is combined with the observed time dilation of a factor $\sim 2$, this ‘energy correction’
implies much higher redshifts of $z \sim 5.5$ (see Figure 8 of Fenimore et al. (1995), which
gives the distance vs time dilation) for the dimmest BATSE bursts, instead of red-
shifts 1 or 2 as reported by Norris et al. (1994, 1995). Clearly, the intrinsic differences
between the duration of GRBs at high and low energies have to be taken into account
when measuring time dilations.

These high redshifts of faint GRBs were at odds with the lower redshifts derived
from the interpretation of the log N–log P distribution, which led to redshifts $z \sim 1$
for the dimmest bursts (Fenimore et al. 1993a, Pendleton et al. 1996b). Mallozzi,
Pendleton, and Paciesas (1996) have used a range of observed burst spectra to analyze
the effects of the spectral shape of GRBs on the cosmological interpretation of the
number–intensity distribution. The spectral parameters considered in the construction
of the cosmological number–intensity distribution has a significant effect on the
value of the maximum redshift derived from the log N–log P models, as much as a
factor of 2. So they gave an average value of the redshift for the faintest detectable
bursts between 0.85 and 1.37 (Mallozzi, Pendleton, and Paciesas 1996). Previous
results which used the BATSE peak flux distribution log N–log P have indicated that
this distribution was consistent at high fluxes with a homogeneous distribution of
sources embedded in Euclidean space (Meegan et al. 1992), and at low fluxes
with non-evolving mono-luminous bursts observed out to redshift $z \sim 1$ (Mao &

How can such large differences in redshift for the dimmest bursts obtained with
time dilation measurements and with the log N–log P distribution be reconciled? One
possibility would be that the observed $-3/2$ power law in the log N–log P distribution
is a coincidence rather than an indication that bright bursts are from a nearby,
homogeneous population.\textsuperscript{2} If, on the other hand, the log $N$–log $P$ distribution indicates a valid distance scale, then a large fraction of the observed time dilation would be intrinsic to the bursts. As log $N$–log $P$ is compatible with a stretching of $\sim1.3$ rather than 2, 65\% of the time stretching would be intrinsic (Fenimore & Bloom 1995). In this case the time dilation is not a good choice to prove GRBs are at cosmological distances.

In a way, these analyses confirmed many doubts raised previously: the time dilation can be intrinsic to GRBs and it is not proven that peak intensity indicates distances. All these discussions explain why many authors remained so skeptical with respect to the time dilation test. Let us now see if the spectral softening of distant GRBs, the dependence of $E_p$ on the burst intensity, provides a better way of proving that GRBs are at cosmological distances.

### 2.4.2 Spectral softening of dim bursts

The $\nu F_\nu$ spectra of GRBs usually exhibit a peak at an energy $E_p$ in the range 20–2000 keV. If GRBs are at cosmological distances the spectra of dim bursts should be redshifted to lower energies relative to those of bright bursts. Mallozzi et al. (1995) obtained photon energy spectra for 400 bursts. The photon spectra used for the determination of $E_p$ were derived in the approximate total duration interval of each event. Of course this determined an average value of $E_p$ because $E_p$ has a significant evolution among many bursts (e.g. Ford et al. 1995). GRBs were considered in five intensity groups. There is a clear increase of $E_p$ with intensity (their Figure 2). Taking a redshift $z_p$ associated with an observed peak photon flux $P$ the maximum range of $z$ consistent with the analyzed sample for peak fluxes between 1 to 100 photons cm$^{-2}$ s$^{-1}$ is given by $(1 + z_1)/(1 + z_{100}) = 1.86 + 0.36$ or $-0.24$. Using $z_{100} \sim 0.1$ found by Mallozzi, Pendleton, and Paciesas (1996) from model fits of the differential number–intensity distribution, the value of $z_1$ is $\sim 1.05 \pm 0.39$ or $-0.27$. The actual redshift of the dimmest bursts was expected to be larger than 1 because the sample has been limited to peak flux greater than $\sim1$ photon cm$^{-2}$ s$^{-1}$ while the dimmest GRBs observed by BATSE have peak fluxes of $\sim0.3$ photon cm$^{-2}$ s$^{-1}$ (Fishman et al. 1994). These values are in closer agreement with redshift estimates derived from the modeling of the number–intensity distribution (see Section 2.4.1). Nevertheless, it has to be noted that a factor of 2 in $z_{\text{max}}$ is possible (Wickramasinghe et al. 1993). Moreover, a wide range of intrinsic luminosities, the choice of the burst spectral index, and of course the characteristics of the cosmological model have a significant effect on the value of $z_{\text{max}}$ (Emslie & Horack 1994, Hakkila et al. 1996).

In conclusion, the origin of the observed time dilation and spectral softening of dim bursts remained controversial and the cosmological interpretation of these effects did not permit deriving secure redshifts for the faintest GRBs detected with BATSE. As noted by Mallozzi et al. (1995), their results are consistent with, but not proof of, sources of cosmological origin. In fact, to prove that GRBs are really at cosmological distances it would be important to demonstrate that time dilation and energy shifting

\textsuperscript{2} We will see in Chapter 3 that this is indeed the case.
measurements agree with other cosmological signatures such as the brightness distribution, in the context of cosmological models. In the BATSE decade, these studies were made very difficult by the lack of knowledge of the intrinsic (luminosity function, correlations between their various spectral and temporal properties) and statistical (evolution) properties of GRBs. Let us see now if gravitational lensing can provide more convincing evidence in favor of cosmological distances.

### 2.4.3 Lensing effects

If GRBs are at cosmological distances they can be lensed by intervening galaxies (Paczyński 1987). A lensed GRB would be characterized by the detection at different times of bursts with similar time profiles and identical spectra coming from the same location. As the detected bursts are noisy it may be difficult to identify two bursts as lensed images. Moreover, the estimate of the frequency of lensed GRBs by Grossman and Nowak (1994) is low: one lensed GRB every 1.5–25 years with a median time delay of about 7 days between images. This rate did not give any great hope of seeing lensed GRBs and in fact no such events have been reported (Nemiroff et al. 1993, 1994c).

It is also interesting to note that gravitational lensing by objects in the mass range \(\sim 10^{17} - 10^{20}\) g (femtolensing) has been studied (Stanek, Paczyński, & Goodman 1993). If dark matter is made of \(\sim 10^{19}\) g objects that contribute significantly to the mass density of the Universe, the probability of such an effect is about the same as the probability of microlensing (Mao 1993). Stanek, Paczyński, and Goodman (1993) have shown that femtolensing could introduce complicated interference patterns that might be interpreted as absorption or emission lines in GRB spectra. Amusingly, the spectral ‘lines’ generated by femtolensing interference patterns, are equally spaced in energy just like the cyclotron lines reported in Chapter 1.

At the end of this discussion on the search for cosmological effects in BATSE GRBs, it is reasonable to say that all these tests have strengthened the cosmological models without providing an answer to this still pressing question: Where are GRBs coming from?

Perhaps the most convincing way of elucidating the mystery of GRB origin would be the discovery of GRB counterparts. As we have seen this search started well before BATSE, but without conclusive results. Let us have a look at the new results obtained during the BATSE decade.

### 2.5 GRB COUNTERPARTS

During the BATSE decade, the searches for counterparts using the GRBs with the best IPN localizations did not find any credible counterpart to a GRB. Schaefer (1994) and Hudec (1993), who were actively involved in these counterpart searches, gave complete reviews on this topic.
In 1993, the failure of the on-board tape recorders of CGRO led NASA to take the decision to transmit all the data from the satellite in almost real-time with the TDRSS communication satellites. This provided the opportunity to calculate the GRB positions on the ground a few seconds after the detection of the burst with the satellite. This opportunity was realized by Scott Barthelmy who set up a burst localization system using near real-time BATSE data. He called this system BACODINE for BATSE COordinates DIstribution NETwork (Barthelmy et al. 1994). BACODINE could distribute GRB positions to external sites within about 5 s from the time of their detection by BATSE. This new system spurred the development of a new generation of rapidly slewing optical telescopes, which could observe the error boxes defined by BATSE within seconds of an alert, an unprecedentedly short delay after the burst. Of course, many difficulties existed, such as the scarcity of such telescopes, their limited sensitivity, the size of the error boxes, which varied from a few degrees to 10 degrees for the weakest bursts, and the visibility of GRBs (which had to be above the horizon, at night, without clouds). At that time other GRB experiments, like Watch on the Soviet GRANAT spacecraft (Brandt, Lund, & Castro-Tirado 1994) and Watch on the European Space Module Eureca (Castro-Tirado et al. 1994b) were able to give more accurate burst positions (several tens of arcminutes) as well as COMPTEL and EGRET on CGRO. Nevertheless none of them could compete with BATSE and BACODINE to give very fast alerts a few seconds after the burst.

In any case all these instruments could deliver GRB error boxes much more rapidly than in the past using the Interplanetary Network. In the 1980s the searches were done months and more often years after the bursts themselves and it was impossible to detect short-lived counterparts. With the third IPN of GRB detectors (Hurley et al. 1993), the delay between burst localization and counterpart searches was considerably reduced. For instance, after GRB 920501 the optical, radio and X-ray counterpart searches started 12, 13 and 19 days after the burst. X-ray and radio sources were found in the error box, with some similarity with the 1978 November 19 burst discussed in Chapter 1 (Hurley 1994b). But, as for the 1978 November 19 burst, no sure association with an X-ray or radio source was possible. In fact the time delay (3 weeks) was still too long.

Another example of a relatively short delay between the GRB detection and the counterpart searches is the case of the intense GRB930131, which was observed with many instruments, within hours and days of its occurrence (Schaefer et al. 1994b). The international campaign to search for fading counterparts started 6.8 h after the burst. Table 1 of Schaefer et al. (1994b) gives the impressive list of observations that were undertaken, unfortunately without finding any fading counterpart. Schmidt telescopes have also been used. For instance, in Europe new facilities were supported by an European Consortium as part of the European Southern Observatory (ESO). A program had been running since 1991 at ESO with the 1-m Schmidt telescope. Its field of view of $5^\circ \times 5^\circ$ was well adapted for sensitive searches inside the BATSE error

---

3 Fortunately, CGRO was equipped with the devices needed to communicate in near real-time with the TDRSS communication satellites.
boxes (Boër et al. 1994c). These authors reported lower limits to the magnitudes of many optical counterparts, with time lags always larger and often much larger than 1 day, the minimum lag being 2 days for GRB 920311, GRB 920517, GRB 920830, and GRB 930118. The planned offset of few hours was not reached, and the best reported magnitude limits were 20 in the B band, and 22 in the R band. Other rapid response searches for optical counterparts were reported by Barthelmy, Palmer, and Schaefer (1994), Castro-Tirado et al. (1994a), Hurley et al. (1994b), and Schaefer et al. (1994).

In the radio, observations were undertaken within few days of the burst. The most extensive study was due to Frail et al. (1994) with GRB 940301. Daily images of the Comptel error box were obtained between 3 and 15 days after the burst. Three deep observations were also performed 26, 47, and 99 days after the burst, with no result. An upper limit of 3.5 mJy was placed on any fading/flaring radio counterpart at 1.4 GHz and 55 mJy at 0.4 GHz. This was the first opportunity to test the predictions of relativistic fireball models on the appearance of time-variable radio emission due to the fireball. Systematic searches for millimetric emission (31–90 GHz) from GRBs in the 3B Catalog of BATSE were done by Schaefer et al. (1997). They used the two-year Differential Microwave Radiometer data from COBE (Cosmic Background Explorer). No significant signal was found before or after the 81 GRBs observed, but the 95% confidence level upper limits were too high (>175 Jy or \(\sim 9.6 \times 10^{-13} \text{erg cm}^{-2} \text{s}^{-1}\) at 31 GHz). A deep near-infrared search for quiescent counterparts of nine GRBs within the smallest error boxes obtained with the 3rd IPN was made by Larson and McLean (1997). \(J\), \(H\), and \(K\) images were obtained with limiting magnitudes: \(J \sim 20.0\), \(H \sim 19.1\), and \(K \sim 18.7\), for 5\(\sigma\) detection of a point source. This was between 5 and 200 times fainter than previous near-infrared searches (Schaefer et al. 1987). No counterpart was reported suggesting that the true host galaxies were fainter.

The first searches for GRB quiescent counterparts in the extreme UV were undertaken with EUVE satellite (Extreme UV Explorer) launched in 1992. The authors aimed at constraining galactic candidates like thermally radiating quiescent counterparts such as nearby hot NS (Hurley et al. 1995). No quiescent emission was detected in a 40 ks long observation of GRB 920325, but the observation was performed 17 months after the GRB. The Hubble Space Telescope (HST) allowed the first deep ultraviolet searches for quiescent counterparts of five GRBs to be made, but again without success (Schaefer et al. 1997).

In X-rays, several small error boxes obtained with the IPN were observed with Rosat (Hurley et al. 1999) or ASCA (Shibata et al. 1997) a long time after the bursts, without success. The fastest X-ray observation was achieved by Boër et al. (1993), using a 2680-s Rosat target of opportunity observation performed 19 days after GRB 920501.

A complete review of past and current counterpart searches in the soft X-ray, UV, optical, infrared and radio bands is given by Schaefer (1994), with an extensive bibliography. All these negative results were analysed by McNamara, Harrison, and Williams (1995). They made some recommendations for future searches. As GRBs do not produce optical counterparts brighter than 10th magnitude from a minute to a day, and brighter than 17th magnitude from 1 to 99 days after the burst, they
recommended for the future to do deep imaging (>12 mag) within 10 hours. From the observations of Frail et al. (1994) reported above, McNamara, Harrison, and Williams (1995) concluded that GRBs do not produce radio counterparts brighter than 10^{-17} \text{erg cm}^{-2} \text{s}^{-1} from 1 to 99 days after the burst. They suggested that radio searches have to extend below 3 mJy and begin within 3 days of the burst detection. In X-rays they mentioned that the delay of the observations has to be considerably reduced, if possible to less than 1 day. In addition to these directions for future X-ray, optical, and radio follow-up observations, McNamara, Harrison, and Williams (1995) made reference to HETE (to be launched in 1996) as a decisive mission for progress in these counterpart searches. The launch of HETE in 1996 would fail, but the same year BeppoSAX would be launched, enabling for the first time in the history of GRBs, the quick determination of accurate GRB positions and the observations suggested by McNamara, Harrison, and Williams (1995).

As it became more and more obvious that counterpart searches performed days or months after the burst were unsuccessful, various observers tried to take advantage of the rapidity of BACODINE. But in general BATSE GRB error boxes were larger than the field of view of the major telescopes in space or on the ground. The desperate need to find the GRB counterparts led to the development of a new generation of optical telescopes with large fields of view, capable of slewing automatically and rapidly in the direction of recently detected GRBs. If the optical afterglow were bright enough, these telescopes would detect them and determine largely reduced error boxes giving the possibility of rapid follow up with large telescopes. In spite of the progress made with the 3rd Interplanetary Network to produce precise GRB locations with reduced time delays, the typical response time was still of the order of half a day (Hurley et al. 1995). The strategy pursued with BACODINE thus appeared the most promising to reach a major goal: the near real-time search for GRB counterparts at many wavelengths. But for the GRB hunt to be efficient the presence of several slewing telescopes around the world was crucial. This strategy of fast searches for GRB counterparts, which appeared so decisive to solve the problem of GRB distances, explains the success of these small telescopes and their developments in different countries. In the US, a sensitive wide-field CCD camera (the ETC for Explosive Transient Camera) started to operate in 1991 for over 3 years at Kitt Peak (Vanderspek, Krimm, & Ricker 1994, Krimm, Vanderspek, & Ricker 1994). Insufficiently sensitive, it never found any convincing evidence for optical flashes due to gamma-ray bursts. A successor appeared, the Gamma Ray Optical Counterpart Search Experiment (GROCSE; Akerlof et al. 1994), which was able to point in the direction of BACODINE alerts, but as its sensitivity was not very much better than ETC (7th or 8th magnitude for 1-s flash) it had not detected any burst after 2 years of operation (Park et al. 1997). In France, Tarot, a 25-cm fully robotic telescope was developed after 1996 (Boër et al. 1999). In the US, efforts to increase the sensitivity of GRB instruments were pursued. At Los Alamos, ROTSE, the Robotic Optical Transient Search Experiment, developed by K. Akerlof, was in operation in 1998 (Marshall et al. 1997) and, at Livermore, LOTIS (Livermore Optical Transient Imaging System) was ready by the end of 1996 (Park et al. 1998). We will return to the role of this new generation of robotic telescopes in the next chapters devoted to
satellites that routinely provided rapid and accurate GRB localizations ideal for these instruments.

After these discouraging results of counterpart searches, we conclude with the main result of the BATSE decade: the growing evidence of the cosmological distances of GRBs. The BATSE observations raised many important issues which are summarized in the next section.

2.6 BEYOND THE DISTANCE PROBLEM, SOME IMPORTANT OPEN ISSUES

In the absence of counterparts and of clear indications of their redshift, the issue of the distance of GRB sources remained open during the early 1990s. Nevertheless a lot of work could be undertaken to solve serious problems that existed for both extended galactic halo and cosmological models. These problems concern the radiation processes and the sources of GRB energy. They were thoroughly analyzed by Harding (1994a,b), and we base the following analysis on the main points of her paper.

The first point is the very high optical depth, $\tau_{\gamma\gamma}$, to photon–photon pair production, which cannot be avoided. The large burst luminosity $L$ and the small size of the source, inferred from the very short timescales observed in the emission fluctuations of GRBs ($R \sim c\Delta t \sim 10^7$ cm) require high isotropic photon densities at the source which lead to huge values of the optical depth to photon–photon pair production: $\tau_{\gamma\gamma} \propto L/R$. For sources in an extended halo as $d \sim 100$ kpc and $L \sim 10^{43}$ erg s$^{-1}$, $\tau_{\gamma\gamma} \sim 10^6$; for sources at cosmological distances $d \sim 100$ Mpc and $L \sim 10^{49}$ erg s$^{-1}$, $\tau_{\gamma\gamma} \sim 10^{12}$.

Such extreme optical depths unavoidably result in a thermalized pair plasma emitting a black-body spectrum at a temperature $kT \sim m_e c^2$. Therefore a turnover or cutoff in the GRB spectrum around 1 MeV is expected, but it is absent in most bursts. The lack of this cutoff in GRB spectra implies that $\tau_{\gamma\gamma} < 1$. Goodman (1986) noticed that an optically thick electron–positron pair plasma at temperature $kT \sim m_e c^2$ would expand adiabatically as a relativistic fluid. This expanding fireball cools as thermal energy is converted into bulk kinetic energy. When it becomes optically thin the radiation can escape. At that time the radiation has a black-body spectrum with $kT_{\text{esc}} \sim 20$ keV, but it is blueshifted in the observer’s frame to $h\nu_{\text{obs}} \sim \Gamma T_{\text{esc}} = T_0$, where $T_0$ is the initial temperature of the fireball. So the escaping radiation has a spectrum that is a modified black-body at a temperature of a few megaelectronvolts. Here again this is not the typical shape of GRB spectra. To avoid these problems it can be assumed that the radiation field is not isotropic, in which case $\tau_{\gamma\gamma}$ can be greatly reduced. Indeed, the threshold for a photon of energy $\epsilon_1$ to produce a pair with another photon $\epsilon_2$ is: $\epsilon_1 \epsilon_2 = 2m_e^2 c^4/(1 - \cos \theta_{12})$; where $\theta_{12}$ is the angle between the two photons. If this angle is very small $\epsilon_1$ can be very high, increasing greatly the threshold for pair production and reducing considerably $\tau_{\gamma\gamma}$.

In a similar way, this large value of $\tau_{\gamma\gamma}$ can also be reduced if a relativistic motion of the source is considered. This motion produces beamed radiation; the emission is beamed within an angle $\theta \sim 1/\Gamma$, where $\Gamma$ is the bulk Lorentz factor of the radiating
matter. $\tau_{\gamma\gamma}$ is reduced because both the required photon density and the photon energy in the comoving frame are reduced. The value of $\Gamma$ which is required to have $\tau_{\gamma\gamma} < 1$ at energy $E_{\text{max}}$ (the maximum energy of the detected photons) has been calculated by Baring and Harding (1993). For example, Harding (1994a) calculated the minimum bulk Lorentz factor $\Gamma$ allowing the escape of GeV photons from GRB sources with typical dimensions fixed by the observed variations of GRB flux: $R = \Gamma c \Delta t = \Gamma R_v$ with $R_v = 10^7$ cm. She considered three bursts: GRB 930131 with $E_{\text{max}} = 1$ GeV, GRB 910814 with $E_{\text{max}} = 60$ MeV, and GRB 910503 with $E_{\text{max}} = 170$ MeV. If these GRBs are assumed to be 30 kpc (in an extended galactic halo model), she finds $\Gamma \sim 10$–13. If the sources are at 1 Gpc (cosmological model), $\Gamma$ is a few hundred, between 240 and 867. Even if the source is very nearby ($\sim 1$ kpc) the bulk Lorentz factor would be $\Gamma \sim 3$–4 (see also Table 3 of Baring and Harding (1997) for a larger number of GRBs). The lower bounds on $\Gamma$ are a little lower but they have the same order of magnitude. Another advantage of relativistic beaming is a smaller luminosity at the source because the observed flux is enhanced by a solid angle factor $\Gamma^2$, $F \propto \Gamma^2 L/d^2$, where $d$ is the distance of the source and $L$ its luminosity (Krolik & Pier 1991). With the above constraints on $\Gamma$ this factor can reach $\sim 10^6$ for cosmological GRBs.

To allow the escape of high-energy gamma-rays Mészáros and Rees (1993b) assume a baryon-loaded fireball (and no longer a pure electron–positron pair fireball). The baryon load of the fireball has to be very limited, $M c^2/E_0 < 10^{-5}$, to avoid losing most of the energy of the fireball into the acceleration of the baryons ($E_0$ is the initial total energy of the fireball and $M$ the rest mass of the baryons, in the following we denote $\eta = E_0/M c^2$). As it expands and cools, the fireball accelerates to $\Gamma \sim \eta$. For these authors the GRB would be produced by internal shocks inside the fireball (see Chapter 5), and the typical values of the bulk Lorentz factor would be $\Gamma \sim 10^2$–$10^3$. When the fireball has swept up enough material, an afterglow would appear due to the shock of the fireball on the external medium. This shock converts the kinetic energy of the fireball into radiation. This shock, and the reverse shock moving back into the fireball, will produce long-lived non-thermal emission, called the afterglow.

The necessity of relativistic beaming in GRB sources, if they are at cosmological distances, was also raised very early by Brainerd (1992a) to explain the non-thermal, rapidly varying spectra of GRBs. Again, the optical depth of $e^-/e^+$ pairs led to Lorentz factors greater than 100. The author also noticed that jet models producing cyclotron lines were allowed. The evolution of cosmic fireballs as GRB sources was already largely discussed by Piran (1994) and Shemi (1994) at the second Huntsville workshop. This is quite normal because the possible models (the extended galactic halo and the cosmological models) faced the same difficulties. As these fireballs are at the heart of GRB origin they will be extensively discussed in Chapters 5 and 6.

These considerations concerning the physical constraints on the plasma emitting the GRB provide no answer to the origin of the sudden large amount of liberated energy. For extended halo models with sources at typically 100 kpc the luminosity is already very high, $10^{42}$–$10^{43}$ erg s$^{-1}$, that is $10^4$ times the Eddington luminosity for a
neutron star. So, with isotropic emission $\tau_{\gamma\gamma}$ already very large, a fireball will form unless the radiation is beamed or the source has a relativistic motion. For these extended galactic halo models, we have seen that high-velocity neutron stars were invoked as a source of energy. These neutron stars may be born in the disk or they can be born in the halo, for instance via white dwarf mergers (Eichler & Silk 1992). In both cases it is not so easy to produce such large amounts of energy. If the emission is highly collimated, the emission will be reduced, but we can see only a small fraction of the existing sources, and it may be difficult to find enough sources in the halo.

In the case of cosmological models, which seem less ad hoc, the problem of the large amount of energy is not trivial but in the end such energies can be found. Two energy sources were proposed at the CGRO epoch.

(a) **Binary NS mergers (more generally the coalescence of two compact objects).** In this scenario the binding energy of the system, which can be released, reaches $10^{53} - 10^{54}$ erg. Most of this energy appears as gravitational waves ($8 \times 10^{52}$ erg) and neutrinos ($2 \times 10^{53}$ erg; Clark & Eardley 1977). Gamma-rays might be produced through the annihilation of neutrinos and antineutrinos ($\nu + \bar{\nu} \rightarrow e^+ + e^- \rightarrow \gamma\gamma$), with an efficiency of $10^{-3}$ which is enough to explain the GRB luminosity.

(b) **Formation of a transient accretion disk around a black hole (Woosley 1993).** This black hole might result from a NS–BH merger or from a failed type Ib supernova. In this last case, if the envelope cannot be liberated, it falls back onto the compact object which has been just formed. Here again the released energy can reach $\sim 10^{50}$ erg.

So, in 1993 the theoretical obstacles for cosmological GRBs were being removed, and the fireball mechanism proposed by Mészáros and Rees (1993a) predicted the existence of long-lived afterglows at longer wavelengths, which remained to be discovered. In the case of extended halos the problem was to introduce a completely new population of compact objects in a new galactic structure: an extended galactic corona. Even if high-velocity neutron stars able to escape from the galactic plane had been observed, introducing a large population of neutron stars in an extended halo seemed much more ad hoc than accepting cosmological GRB sources. Moreover, even with these sources, the energy involved was very large and static source models could not work since the radiation pressure would drive a wind away from the star. A cosmological origin was finally more attractive and supported by different CGRO results: the absence of cyclotron lines, the isotropic and inhomogeneous source distribution, and the possible detection of various cosmological effects (time dilation, spectral softening, number-intensity distribution) with converging $z$ values. At that time, however, it was difficult to go further in the construction of GRB models without knowing the distances of the sources. Clearly, what was needed was the identification of long-lived counterparts. We will not have to wait too long to elucidate the origin of GRBs, the solution will arrive in 1997 as explained in Chapter 3.
2.7 CONCLUSION

BATSE on CGRO had a decisive impact in the GRB story. The breakthroughs were the spatial and angular distribution obtained for the weakest bursts. Galactic disk models were almost unanimously rejected, even if two-component models were still defended (essentially to preserve a very local galactic population which could emit the cyclotron lines). Cosmological models gained ground with important parallel theoretical works on the fireballs. But no decisive observation could provide the distance of the sources. The best proof is the debate that took place at the Baird Auditorium of the Smithsonian Museum of Natural History in Washington, on April 22, 1995. B. Paczyński and D. Lamb were the natural candidates to debate on the GRB distance scale, M. Rees was the chairman and the moderator. V. Trimble introduced the debate with a reminder of the 1920 debate between Curtis and Shapley about ‘The scale of the universe and the nature of spiral nebulae’. B. Paczyński (1995) defended of course a cosmological origin of GRBs, suggesting the coalescence of compact objects as the energy source, and D. Lamb (1995) an extended galactic halo with explosions at the surface of high-velocity neutron stars. This exciting debate did not allow a conclusion to be reached (Rees 1995). The two camps stayed on their positions and the audience was divided. We nevertheless quote the premonitory conclusion of M. Rees: ‘The controversies in the Shapley/Curtis debate were settled within a few years. The subject thereby made a forward leap, and astronomers moved on to address more detailed issues. I’m enough of an optimist to believe that it will only be a few years before we know where (and perhaps even what) the gamma-ray bursts are.’ To progress it was urgent to find the afterglows predicted by fireball models in radio (Paczyński & Rhoads 1993), X-rays (Vietri 1997), and optical wavelength (Mészáros & Rees 1997). This would be wonderfully done thanks to BeppoSAX: our next step and chapter in the GRB saga.

2.8 REFERENCES


3

BeppoSAX and the afterglow era

Before analyzing the major breakthroughs of BeppoSAX, which allowed the first multi-wavelength detections of GRB afterglows, we will say a few words about the mission and its instruments. In the introduction to this chapter, it is important to mention the remarkable work done by the scientific and technical teams, which were strongly motivated by two researchers, E. Costa and F. Frontera. These two researchers were very excited by the GRB mystery and they knew that the set of instruments on-board BeppoSAX were particularly well adapted to the hunt for GRBs and specifically to their rapid follow-up which, as we have seen, had been impossible earlier.

Section 3.2 will be devoted to an analysis of a few major GRBs which, thanks to fast multi-wavelength follow-up, made it possible to fix the GRB distance scale, to detect GRB host galaxies and to support the relativistic fireball model as the most convincing to explain the bursts and their afterglows. For this giant step forward we will dwell on the 1997–1999 period, which corresponds to the last third of BATSE’s life. We will particularly emphasize the year 1997, with three major GRBs which enlightened the domain after a quarter of century of observations which could not elucidate the GRB mystery.

In Section 3.3 we will discuss various observational issues concerning the connection between the GRB and its afterglow, the existence of dark GRBs, the evidence in favor of jetted emission and the prediction of orphan afterglows, the GRB–supernova connection, and the presence of spectral features in the afterglows at high energy. These issues will be analyzed more deeply in Chapter 4 thanks to new results from other very successful missions such as HETE-2, INTEGRAL and Swift. Finally we will introduce the $E_{\text{peak}}-E_{\text{iso}}$ (Amati) relation which can be important in estimating the GRB distances when the redshift is not available and which might be the signature of the physics underlying the prompt GRB emission.

In this chapter, we will discover the temporal and spectral evolution of many GRB afterglows and, at the end of this chapter, it will be clear that the solution of the
GRB mystery could only come from the fast and more accurate localizations of BeppoSAX. Before this mission afterglow searches were either fast but insufficiently accurate (with BATSE and BACODINE) or accurate but not fast enough (with the IPN). In reality some IPN localizations were of sufficient quality for afterflow searches since 13 of them resulted in the discovery of an afterglow in the period 1999–2006 (including GRB 000131 at \(z = 4.5\)), when we knew what to search.

### 3.1 BEPPOSAX: ITS INSTRUMENTS, ITS STRATEGY

The BeppoSAX satellite (SAX for Satellite italiano per Astronomia X, and Beppo from the nickname of the famous Italian physicist, Giuseppe Occhialini), launched from Cape Canaveral on April 30, 1996 was put into an almost equatorial orbit \((i = 3.9^\circ)\) at an altitude of 600 km. This orbit is well protected from high-energy cosmic rays by the Earth’s magnetic field; so the background of X-ray and gamma-ray detectors due to these high-energy particles is minimized.

The GRB detection on BeppoSAX was based on the GRB Monitor (GRBM), composed of four identical CsI(Na) scintillator slabs, about 1000 cm\(^2\) each. This system, with a nearly 4\(\pi\) field of view, worked between 40 and 700 keV (Frontera et al. 1997, Costa et al. 1998). The localization was provided by two Wide Field Cameras (WFC; Jager et al. 1997), which were position-sensitive proportional counters with a coded mask in front of their collimated field of view. The WFCs had an energy range of 2–26 keV and a field of view of \(40^\circ \times 40^\circ\) at zero response. After the detection of a GRB by the GRBM, the scientific team looked for the burst in the WFCs, and if the burst was visible in a WFC it could be localized with a precision of 3 arcminutes. When the GRBM triggered on a genuine GRB, it could be from any place on the sky, with a 10% probability of being in the field of view of one of the WFCs, given the field of view of these cameras. This permitted the localization of about 10 GRBs per year, but these events revolutionized the field. The combination of a triggering and an imaging instrument was an original approach to solve the problem of providing fast, accurate GRB positions. Since the WFCs had no trigger capability and could not measure source positions in-flight, this job had to be done on the ground. However, in the energy range of the WFC, galactic X-ray bursts are about 100 times more frequent than gamma-ray bursts. Under these conditions the triggers from the GRBM were crucial to alert the scientists to search for small count rate excesses in the WFC at the times of gamma-ray bursts (with about one chance in 10 of success) and, when an excess was found, to construct the image of the sky at that time, to verify that the localization was not that of a known X-ray source, and to quickly distribute the position of the new source to the community of GRB observers. After few years of this work, J. Heise and his colleagues had enough experience to identify 17 transient events in the WFC, which were very similar to X-ray counterparts of gamma-ray bursts, but were not associated with known X-ray sources or triggers in the GRBM (Heise et al. 2001). This would result in the discovery of a class of soft GRBs, which are not detected at gamma-ray energies, called the ‘X-ray flashes’, discussed in Section 3.8. The capability of BeppoSAX to obtain arcminute
positions in a few hours was the major advantage of BeppoSAX over previous missions and the key to its success. This success was due in large part to the use of standard X-ray localization techniques and to the dedication of the team of duty scientists who carefully looked at any tiny excess in the WFC data around the times of GRBM triggers.

In addition to the fast localization capability provided by BeppoSAX, it was possible to quickly reorient the spacecraft to observe GRBs with the narrow field instruments (NFI; Boella et al. 1997b, Manzo et al. 1997, Parmar et al. 1997), which are sensitive telescopes with fields of view of 1.5°. These new pointings could be acquired within 5 to 8 hours of the GRB. After this time, GRB observations in the X-ray domain were possible with the NFI, composed of one low-energy concentrator (0.1–10 keV), three medium energy concentrators (2–10 keV), and one High Pressure Gas Scintillation Proportional Counter (4–120 keV). This complete set of instruments on a single satellite (Boella et al. 1997a), and the capability of reacting quickly (in a few hours), were decisive advantages over previous missions. Since the positions obtained with the WFC were immediately communicated to the scientific community, this allowed the activation of space and ground telescopes working at other wavelengths a few hours after GRB detection. This remarkable strategy was not so easy to impose on the BeppoSAX scientific community at large, which was more concerned with the observations of galactic or extragalactic X-ray sources, and which nevertheless had to accept sudden changes in the observing program to follow some possible evanescent X-ray emission from these mysterious GRBs.

This strategy was nevertheless the right one. It resulted in the discovery of the X-ray afterglows of many GRBs, and in follow-up observations of these afterglows in the visible, radio, and infrared domains. As we shall see with a few examples, this resulted in the determination of GRB redshifts and in the identification of GRB host galaxies. These results also made it possible to understand why it had been impossible to find the afterglows of GRBs in the past, and to give the requirements for future successful GRB missions such as Swift. We illustrate below the contribution of BeppoSAX to GRB science with the presentation of some ‘milestone’ GRBs. A more detailed discussion of GRB afterglows is given in the following sections.

3.2 SOME GRBS SEEN IN 1997 AND THEIR AFTERGLOWS

3.2.1 GRB 970228, the first X-ray afterglow and optical counterpart

Before the detection of this burst a first opportunity to react quickly and to produce arcminute-size WFC error boxes was tested on January 11, 1997. A 10′ radius WFC error box was obtained (Costa et al. 1997c). The satellite was rapidly pointed with its narrow field instruments (NFI) in the direction of the GRB. Several faint X-ray sources were detected with the NFI. However, when three weeks later an improved (3′ radius) WFC error box was announced (in’t Zand et al. 1997) none of the previous
objects were inside this new box. The optical observations which started less than 1 day after the burst (a very fast reaction indeed) gave no evidence for a possible optical counterpart (Gorosabel et al. 1998b).

The next opportunity, GRB 970228, was the good one. This GRB was detected by the GRBM on February 28, 1997 at 2h58 UT. This was a moderately intense, multipeak, classical GRB (Costa et al. 1997a). One of the two WFCs localized this GRB with a precision of 3 arcmin. The decision to look at this burst with the narrow field instruments (NFI) was quickly made and the observation started 8 h after the trigger. The first observation, which lasted about 14 000 s, identified a previously unknown X-ray source with a flux of $3 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$ in the 2–10 keV energy range. During the observation, it appeared that the source was fading. A new observation (lasting 16 000 s) was planned for March 3. Figure 3.1 shows the two images of the GRB region. Clearly the X-ray source had a very fast-decreasing intensity. The decrease followed a power law, with $F(t) \propto t^{-1.3\pm0.1}$ (Costa et al. 1997b), as shown in Figure 3.2. This was the very first detection of an X-ray afterglow following a GRB. One week later ASCA observed the source (Yoshida et al. 1997) and confirmed that the afterglow was continuing to decay like $t^{-1.3}$. The X-ray satellite Rosat also detected this afterglow at a flux level compatible with the extrapolation of the decay law established by BeppoSAX, and provided an improved
Figure 3.2. Variation with time of the X-ray flux from GRB 970228, in the 2–10 keV energy range. Data points from TOO2 (at $4 \times 10^5$ s) are grouped into one point due to lower statistics. Data are fitted by a power law ($F \propto t^{-1.32}$) shown as the solid line at the lower right of the figure. The extrapolation of this law is consistent with the flux detected by ASCA on March 7.028 ($7 \times 10^5$ s; Yoshida et al. 1997). When the same law is extrapolated backwards (dotted line) to the time of the GRB, it gives a good match to the average flux detected by the WFC during the last part of the GRB (the three minor pulses of the time profile, from 35 to 70 s; see Costa et al. 1997e).

position for the X-ray source (Frontera et al. 1998). The extrapolation of this power law backward, to the time of the burst, was in agreement with the flux reported by the WFC for the three last pulses of the GRB light-curve (Figure 3.2, from Costa et al. 1997e). Finally, the BeppoSAX localization was confirmed by the IPN, which included the Ulysses spacecraft far from Earth (Hurley et al. 1997).

In parallel with the X-ray observations, an observational campaign started, using optical telescopes. An optical transient was discovered (Groot et al. 1997b, van Paradijs et al. 1997) thanks to the William Herschel Telescope and the Isaac Newton Telescope (Figure 3.3). The observations took place the night of the GRB appearance and one week later; the afterglow was detected in both observations. Between these two detections the flux had decreased dramatically, by a factor of 300. This was the first true detection of an optical afterglow following a GRB. Subsequent images with the ESO New Technology Telescope (NTT; Groot et al. 1997a) and the Keck Telescope (Metzger et al. 1997a) showed an extended object (with an angular size of $\sim 1''$) at the location of the optical transient (OT), likely the host galaxy. During this period the excitement was at its maximum and two pointings of the HST (Hubble Space Telescope) were planned for late March and early April 1997 (Sahu et al. 1997a,b), and in September 1997 (Fruchter et al. 1997). The point source whose
brightness decayed according to a power law was detected near the edge of a nebulosity, which was interpreted as the host galaxy of the GRB. Much later, new observations confirmed the presence of the host galaxy and established its cosmological redshift as $z = 0.695$ (Djorgovski et al. 1999, Bloom, Djorgovski, & Kulkarni 2001), proving the cosmological distance of this GRB. But, as we will see in the next paragraph, this was not the first redshift determination.

Figure 3.4 (upper panel) gives the light-curves of this burst from X-rays to near infrared, clearly showing the power law decline. This decline compares quite well with the predictions of the simplest fireball model invoking a relativistic blast wave that radiates its energy when it decelerates in the surrounding medium (from Wijers, Rees, & Mészáros 1997). Beyond a few days the light-curves present a bump, which might indicate the contribution from an underlying supernova (Figure 3.4, lower panel, from Galama et al. 2000; see also Reichart 1999).

As a conclusion we can say that this single burst very rapidly made it possible to detect the first X-ray and optical afterglows (no radio afterglow was observed, Frail et al. 1998), and identify the first host galaxy and measure its redshift (but historically the first redshift would be obtained for GRB 970508). These discoveries strongly supported the fireball model (see Wijers, Rees, & Mészáros 1997), possibly in a stellar wind environment (Chevalier & Li 1999).
Figure 3.4. (a) Light-curves of the afterglow of GRB 970228 at various wavelengths, from X-rays to near-infrared. The lines indicate a quite convincing comparison with the predictions of the simplest external shock scenario (Wijers, Rees, & Meszaros 1997). (b) The afterglow of GRB 970228 measured in the V, R_c, and I_c bands at late times (beyond 1 day; Galama et al. 2000—Figure 3.3). The light-curves present a bump that has been interpreted as the contribution from an underlying supernova (Reichart 1999, Galama et al. 2000). The afterglow is modeled with a power-law decay ($F \propto t^{-1.51}$) plus the contribution of a supernova like SN 1998bw, redshifted at $z = 0.695$, the redshift of GRB970228 (Djorgovski et al. 1999).
3.2.2 GRB 970508, the first distance and the first radio afterglow

A new step forward was taken with this burst. It is the first GRB whose distance was measured. After the detection of the burst by the GRBM and its localization by one WFC on May 8, 1997 (Costa et al. 1997d, Heise et al. 1997), a follow-up observation with the narrow field instruments of BeppoSAX was quickly planned; it started 5.7 h after the trigger. The X-ray and optical afterglows were identified, but their light-curves were more complex than in the case of GRB 970228 (Pedersen et al. 1998; Figures 3.5 and 3.6). During the first interval, from 3 to 8 h after the burst the optical brightness was constant or slowly declining. Later the brightness started to increase rapidly and reached a maximum 40 h after the burst. After this maximum, the GRB brightness declined following a power law \( F \propto t^{-1.21} \). Observations taken 96 days later, found a source with \( m_R = 24.28 \pm 0.10 \), much brighter than the extrapolation of the power law (Figure 3.5, from Pedersen et al. 1998); this was the host galaxy.

The X-ray light-curve (2–10 keV) of the GRB and its afterglow were measured with BeppoSAX (Piro et al. 1998a; Figure 3.6). The best fit for the X-ray light-curve is

![GRB 970508 - Optical Transient](image)

**Figure 3.5.** R-band light-curve of the optical transient associated with GRB 970508. Following a period of modest decline, a peak of optical emission was reached about 40 h after the GRB. After this peak, the emission declines following a power law \( F \propto t^{-1.21} \). A possible contribution of a constant source, the host galaxy with \( R \)-magnitude of 25.5 is shown by the curved line at the lower right of the figure (from Pedersen et al. 1998; see also Figure 3 of Fruchter et al. 2000b).
Figure 3.6. X-ray light-curve (2–10 keV) of GRB 970508 and its afterglow (WFC and NF1). The dashed line shows the best-fit power law. The lower panel displays the behavior of the source in the optical R band, from $10^4$ to $10^6$ s (see also Figure 3.5), clearly showing a late time ($\sim$1 day) outburst that is concomitant with a bump in the X-ray light-curve (from Piro et al. 1998a).

a power law $t^{-\alpha}$ with $\alpha = 1.1 \pm 0.1$, compatible with the slope measured in the optical. From $6 \times 10^4$ s to $5.8 \times 10^5$ s after the GRB (17 h to 160 h), a bump appeared in the afterglow light-curve. The energy excess in this bump is $\sim$5% that of the GRB. This bump roughly corresponds in time with the bump seen in the optical light-curve (Figure 3.6; Piro et al. 1998a), suggesting a common origin of both events. The energy released in the 2–10 keV range integrated from 27 s to $5.8 \times 10^5$ s corresponds to about 20% of the total energy released by the GRB in X-rays and $\gamma$-rays.

Piro et al. (1999) reported a possible iron line at 3.5 keV in the soft X-ray spectrum of the early part of the afterglow (the first 16 h). The line disappeared 1 day after the burst, at about the time at which the X-ray outburst appeared at $\sim 6 \times 10^4$ s. These observations were interpreted as implying that the site of the burst was embedded in a large mass of material, possibly the pre-explosion ejecta of a very massive star (Piro et al. 1999). Yoshida et al. (1999) reported the observation of a similar re-burst and of a possible line feature in GRB 970828 observed by ASCA. Again, these results were taken as indications of a link between GRBs and massive star explosions as we will see later.
What was very important for this burst was the brightness of the optical counterpart, which remained high for more than two days, permitting measurement of its spectrum with the Keck Telescope (Metzger et al. 1997b). This spectrum showed absorption lines of FeII, MgI and MgII at redshifts $z = 0.767$ and $z = 0.835$, indicating that the GRB was at or beyond the larger redshift. GRB 970508 occurred 7 billion years in the past (with the currently accepted cosmological parameters), when the cosmos was $1/(1 + z) = 54\%$ of its present size. This measurement marked the end of a 24-year-long debate on the distance scale of GRBs and on the amount of energy released in these events. Finally the host galaxy was identified with the Keck Telescope (Bloom et al. 1998a) and the HST (Fruchter et al. 2000b) when the afterglow had faded below detection, several months after the GRB. The afterglow was found to be exactly coincident with the center of its host galaxy (Fruchter et al. 2000b) leading to the conclusion that the GRB was within the galaxy and not behind it. The spectrum of the galaxy exhibited emission lines which were identified with lines from OII and NeIII at a redshift $z = 0.835$ (Bloom et al. 1998a). It is a blue actively star-forming dwarf galaxy (Fruchter et al. 2000b, Natarajan et al. 1997, Pian et al. 1998).

Another novelty for this GRB was quite important and different from the GRB 970228. It was the first GRB for which a radio afterglow was detected (Frail et al. 1997). During the first month the average radio flux was $\sim 0.6\,$mJy with irregular variations due to interstellar scintillation in our Galaxy, i.e. multiple scattering of ionized galactic gas (like the twinkling of stars in the Earth’s atmosphere). The damping of the fluctuations indicated the increase of the size of the radio emitter. Frail et al. (1997) concluded that the radio-emitting region expanded with a velocity very close to the speed of light. For sure, this result supported the relativistic fireball model of GRBs and the interpretation of the afterglows as due to the shock of the fireball on the surrounding medium (Waxman 1997). The afterglow of GRB 970508 was the first to be observed simultaneously over the whole electromagnetic spectrum, several hours after the burst. It has been possible to reconstruct its multi-wavelength spectrum from X-rays to radio (Galama et al. 1998b). In Figure 3.7, three spectral breaks connecting different power law segments appear clearly, at three frequencies: $\nu_a$, $\nu_m$, $\nu_c$. This behavior is fully consistent with the fireball model (see Chapter 6). The complete spectrum with its breaks strongly supports the idea that GRB afterglows are powered by synchrotron emission of electrons accelerated in a relativistic shock.

The large Lorentz factors involved in the fireball model (Chapter 6) led to predictions of afterglow emission at longer wavelengths as the burst ejecta decelerates and interacts with the surrounding material (Paczyński & Rhoads 1993, Katz 1994, Mészáros & Rees 1997). So the characteristic frequency of the afterglow emission depends on the Lorentz factor of the burst remnant, which decreases as the remnant evolves. These predictions have been confirmed by BeppoSAX for the X-ray afterglow and by follow-up observations of optical and radio afterglows. So, quite remarkably, the first two GRBs with an afterglow confirmed two crucial predictions of the fireball model concerning the temporal evolution of the afterglow (Wijers, Rees, & Mészáros 1997), and its broadband spectral shape (Galama et al. 1998b).
3.2.3 The conference on the island of Elba

The importance of discovering the afterglow of GRB 970228 was immediately realized by the GRB community. To share the excitement of this discovery, and to discuss the new perspectives that it opened up, the Italian Space Agency (ASI) quickly organized a small workshop on the wonderful island of Elba on May 26–27, 1997. At the time of the workshop, the existence of afterglows had already been firmly established, with the discovery of the counterparts of GRB 970228, GRB 970402, and GRB 970508, and the distance of this last GRB had just been measured. This was a time full of excitement, with long-lasting discussions on the possible origins of GRBs, and on the best ways to study them in the light of the discovery of their afterglows. The island of Elba in May was an ideal place for this workshop (entitled ‘Latest Developments in Gamma-Ray Bursts’), allowing passionate outdoor discussions late in the night. A brief summary of this workshop, written by C. Kouveliotou, can be found on the internet at the following URL: http://heasarc.nasa.gov/docs/sax/press/elbagrb.html

3.2.4 GRB 971214, a very distant burst

At the end of this famous year 1997, a relatively bright burst (GRB 971214) was detected by BeppoSAX. An optical afterglow was observed (Halpern et al. 1998) in
the 3.9 arcmin radius error box obtained by the WFC. The first optical images were acquired in the near-infrared 3.2 h after the burst with the 3.5-m Calar Alto telescope (Gorosabel et al. 1998a), and later images were obtained in the R, I, and J bands (Diercks et al. 1998, Halpern et al. 1998). Deep I band images (0.8 µm) obtained on 31 December showed a fading of the afterglow candidate by 3.4 mag at least, confirming the identification of the afterglow. The decay in the R, I, and J bands followed a power law ($t^{-\alpha}$) with an index $\alpha = 1.4 \pm 0.2$ similar to the decay rate of the two previously analyzed GRBs (with indexes of 1.1 and 1.2). But the OT associated with this burst was the faintest of the three and it was also reddest. Observations at 8.46 GHz conducted with the VLA for several days after the burst did not result in the identification of a radio counterpart.

Follow-up spectroscopic studies were performed using the Low-Resolution Imaging Spectrograph (LRIS) mounted at the Cassegrain focus of the Keck II 10-m Telescope (Kulkarni et al. 1998a). Extrapolating the power law decay they anticipated $R = 27.4 \pm 0.8$ mag on 1998 January 10. In fact, they found an object with $R = 25.6 \pm 0.17$ mag at the position of the OT. This extended object was considered to be the host galaxy, which had become visible due to the decay of the OT light. The spectroscopic observations made with the LRIS after the OT has faded enabled identification of a prominent emission line at 5382.1 Å in the optical spectrum. The authors proposed to interpret it as a redshifted Ly$\alpha$ emission line, which is observed at $\lambda = 1215.7$ Å in the rest-frame. The calculated redshift of the host galaxy was $z \sim 3.42$, and its spectrum was that of an actively star-forming galaxy. This large distance had three remarkable consequences: (1) GRB 971214 released a fantastic amount of energy, (2) GRBs like GRB 971214 can be detected at even higher redshifts (if they exist there) since GRB 971214 was detected as a rather bright event, and (3) the host galaxy was observed at a time where the universe was only 2 billion years old, about one-seventh of its present age. Using standard cosmological parameters ($H_0 = 65$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_0 = 0.3$), the luminosity distance of the galaxy is $D_L = 9.7 \times 10^{28}$ cm and the isotropic energy released in γ-rays is $4\pi D_L^2 F/(1+z) \sim 3 \times 10^{53}$ erg since the fluence of this burst was about $1.1 \times 10^{-5}$ erg cm$^{-2}$ (Kippen et al. 1997). This huge energy is comparable with the gravitational binding energy released during the merger of two neutron stars, but in this case most of the energy would be released in neutrinos (Narayan, Paczyński, & Piran 1992). It is much larger, by two orders of magnitude, than the typical energy released by supernovae in electromagnetic radiation. But Paczyński (1998) thought that the death of extremely massive stars could produce such large energies and he introduced the word ‘hypernova’ to identify these most extreme events. In this case, the core of a very massive star might collapse directly into a black hole, leaving a dense torus of material. Accretion of the torus onto the black hole would produce plenty of energy. In the same way the merger of two neutron stars would lead to a black hole surrounded by a torus. In both cases, the gravitational collapse produces rapidly spinning objects which might eject their energy in two narrow beams along the axis of rotation. This collimation might reduce by a large factor (a hundred or more) the total amount of energy needed, with respect to the amount calculated in the case of an isotropic emission. Of course we will come
back to these models in Chapter 5, and on the question of jets and beaming in Section 3.4.

Through the multi-wavelength studies of only three GRBs observed in 1997, we have described the major breakthroughs made by BeppoSAX. The distances and the energetics of GRBs were now fixed. Deep optical observations of the afterglows with the largest telescopes have made it possible to study their light-curves, to determine the redshifts and the nature of the host galaxies. Multi-wavelength studies of the afterglows have confirmed that the relativistic fireball model provides the proper theoretical framework to describe how the energy is liberated and propagated. Decidedly, a major step forward was made in the understanding of the GRB mystery. These fundamental observations have also shown the diversity of the afterglows, like the diversity of the GRBs themselves. The optical afterglows, as we have seen, can follow different evolutions (see for instance Figures 3.4 and 3.5) and can even be absent. We will see below that there are many ‘dark GRBs’, with X-ray counterparts but no optical counterpart. In the same way the radio afterglows are often absent— for more than half of the observed GRBs.

3.3 DARK GRBS

Through the previous examples we have seen that GRB afterglows may not be detectable over the whole electromagnetic spectrum. The first well studied bursts, GRB 970228 and GRB 971214, were not visible in radio but both were visible in the optical and in X-rays. Quickly, it appeared that in comparison with the X-ray afterglows which exhibit moderate variations and are present in most GRBs (90%), the optical afterglows are less frequent (50%) and they vary enormously. GRBs which are not visible at optical wavelengths are called dark GRBs.

A first example of a dark burst was GRB 970828, a bright GRB in γ-rays. From 4 h to 8 days after the burst no optical counterpart varying by more than 0.2 magnitude was found, down to a magnitude $R = 23.8$ (Groot et al. 1998a). Comparing the ratio of the optical peak flux to the γ-ray fluence of GRB 970828, with the same ratio for GRB 970508, Groot et al. (1998a) found that the optical peak response of GRB 970828 was at least a factor of $\sim 10^3$ fainter than that of GRB 970508. The standard afterglow model, which worked so well for GRB 970228, predicted an optical flux 300 times greater than observed.

How to explain these dark bursts? In their paper, Groot et al. (1998a) discussed two possibilities: a jet model with an ad hoc beam pattern ($E \propto \theta^{-4}$, and $\Gamma \propto \theta^{-1}$), producing an optically poor afterglow, or the presence of absorbing dust in the host galaxy. They noted the strong influence of the redshift on the estimation of the optical absorption from the low-energy cut-off observed in the X-ray afterglow spectrum. In our galaxy the column density of hydrogen, $N_H$, and the optical extinction (for instance $E(B - V)$) can be related straightforwardly: the usual formula is $N_H/E(B - V) = 4.93 \times 10^{21}$ cm$^{-2}$ mag$^{-1}$ for a standard galactic extinction law— Diplas and Savage (1994) and Predehl and Schmitt (1995). At high redshift the photons observed at a given energy have been emitted at the source with an energy
(1 + z) times greater, with the consequence of having a much higher optical extinction for a given measured \(N_H\). The larger optical extinction is explained by the combination of two effects: first, the cut-off observed in the X-ray spectrum is at higher energy in the rest-frame of the source (implying a true \(N_H\) higher than observed), and second, the photons observed in the visible domain were emitted at the source in the UV domain were the extinction is considerably higher. For a GRB at \(z \sim 1\), and an apparently moderate \(N_H = 10^{21} \text{ atom cm}^{-2}\), Groot et al. (1998a) found that these combined effects would lead to an extinction of \(\sim 5\) mag in the R band.

The lack of optical emission from GRB 970828 was explained later by Djorgovski et al. (2001) as being due to intervening dust. They showed that a single intervening giant molecular cloud would provide the necessary extinction. Another example is provided by GRB 980329, which was one of the brightest gamma-ray bursts detected by WFC. An X-ray afterglow was present (in’t Zand et al. 1998), and a variable radio source was also detected with the VLA (Taylor et al. 1998). A faint optical counterpart was identified at a position consistent with that of the VLA source (Palazzi et al. 1998). Here again extinction by dust in the vicinity of the GRB or in the host galaxy might explain the very faint afterglow of this burst. Finally, GRB 000210, the brightest GRB detected with the GRBM on BeppoSAX was also a dark GRB with X-ray and radio afterglows, but no optical afterglow despite prompt and deep optical follow-up searches \((R > 23.5\) at \(T = T_{\text{trig}} + 16\) h; Piro et al. 2002). Interestingly, in this last case the extinction required to explain the lack of optical afterglow in the context of the fireball model \((A_R = 1.8 \text{ to } 6\) magnitudes) is fully consistent with the extinction inferred from X-ray observations, if one assumes a galactic extinction law at the redshift of the burst, \(z = 0.846\).

An interesting question connected with the inferred dusty environment of some GRBs is the possibility for the bursts to destroy the dust and ionize the gas in their environment. This question was raised after the measure of a high column density of hydrogen associated with a low visible extinction in some GRBs. The impact of GRBs on their environment, and their ability to make themselves visible, was discussed by various authors (e.g. Waxman & Draine 2000, Galama & Wijers 2001, Fruchter et al. 2001) with the conclusions that dust destruction will certainly take place within a few tens of parsecs from the burst, and that it will manifest itself by rapid color changes of the visible afterglow in the minutes following the burst. In the case of GRB 000210, however, the possibility of significant gas ionization by the burst was rejected by the X-ray spectrum which showed no evidence for partially ionized gas on the line of sight to the GRB (Piro et al. 2002). We will come back to these questions in the final chapter (Chapter 9).

Dark GRBs can also be explained by a rapid decay of the optical counterpart (e.g. GRB 980326, Groot et al. 1998b; GRB 021211, Crew et al. 2003). GRB 980326 presented an optical counterpart which decayed very rapidly (Groot et al. 1998b). Its temporal decay was well represented by a power law with an index \(-2.10\), when the typical index is around \(-1.3\). Such a rapid decay could explain the non-detection of some afterglows, when the observations started several hours after the GRB. Similarly, the visible afterglow of GRB 021211 was fainter than \(R \sim 23\) one day after the burst. Other explanations might be linked with the origin of GRBs, for
instance if they are produced in galactic halos by merging neutron stars in a very clean environment (see Chapter 8). In the course of their study of GRB 000630, a GRB with a very faint visible afterglow, Fynbo et al. (2001) provided a list of possible explanations for the lack of detectable visible emission in dark GRBs:

(i) there is a optically bright vs. optically dark dichotomy for GRBs similar to the radio quiet vs. radio loud dichotomy for QSOs (i.e. a bimodal OT luminosity function), (ii) a large fraction of GRBs occur at redshifts \( z > 7 \) and are hence invisible in the optical due to \( \text{Ly}\alpha \) blanketing and absorption by intervening Lyman-limit systems, (iii) a large fraction of GRBs occur in highly obscured galaxies similar to the SCUBA-selected galaxies (e.g. Ivison et al. 2000) so that the optical emission never escapes the host galaxies, or (iv) the shape of the OT luminosity function is such that with the search strategies applied in the period 1997–2000 we would not expect to detect OTs for more than 30\% of the well localized GRBs.

They concluded that the majority (\( \geq 75\% \)) of dark GRBs could be explained by an insufficient depth of the optical observations, if many GRBs have a visible emission like GRB 000630. In other words they could not conclude that GRBs with and without detected OTs belong to two different classes.

Overall, about 40\% of BeppoSAX GRBs had no optical counterpart. The properties of these bursts have been summarized by Piro (2004):

- The X-ray flux of their afterglows is on average a factor of 6 lower than that of GRBs with optical afterglow.
- In 75\% of dark GRBs the upper limit of the optical to X-ray flux ratio \( f_{\text{ox}} \) is consistent with the ratio of GRBs with optical afterglow. This might suggest that many GRBs appear dark due to insufficiently fast or deep visible searches.
- For \( \sim 25\% \) of them \( f_{\text{ox}} \) is at least a factor of 5–10 lower than the average value for GRBs with optical afterglow. They cannot be explained by dim GRBs with afterglow. They seem truly dark and the optical depletion may be due to absorption by dust in star-forming regions. Piro suggested that a fraction of these bursts could be at very high redshift (\( z > 5 \)). The fraction of high-z GRBs (\( z > 5 \)) has, however, been estimated to be less than 20–30\% by Bromm and Loeb (2002).

Given these various possibilities, it was necessary to have a criterion to define ‘truly dark’ GRBs. Jakobsson et al. (2004) proposed such a criterion, based on the measure of \( \beta_{\text{ox}} \), the spectral index of the emission interpolated between the optical and the X-ray domains. ‘Truly dark’ GRBs are those for which \( \beta_{\text{ox}} < 0.5 \). This value now provides a standard criterion to identify dark bursts (see also Rol et al. 2005). Hence Jakobsson et al. (2004) show (their Figure 1) that for a sample of 52 GRBs, only five are truly dark.

In conclusion, the study of these dark bursts is interesting to probe the circumstellar matter around the GRB source or the interstellar medium in the host galaxy. We will come back later to this problem of dark bursts, but it was interesting to put
this class of bursts in the long list of the discoveries made with BeppoSAX and its follow-up observations.

3.4 JETS AND COLLIMATION

The distance scale of GRBs being established, the major remaining uncertainty in the determination of the burst energy is the knowledge of the solid angle \( \Omega \) into which the burst radiates. If this solid angle is small with respect to \( 4\pi \) steradians (isotropic emission), the amount of energy radiated by the burst can be reduced by a large factor (equal to \( \Omega/4\pi \)). But at the same time the event rate in the Universe is increased by the inverse factor, \( 4\pi/\Omega \). This phenomenon of collimation is classical in astrophysical sources: protostars, galactic microquasars, active galactic nuclei, and blazars present bipolar outflows or jets, which indicate that the energy and the matter are often emitted anisotropically. The question of GRB collimation was first considered by Rhoads (1997).

The collimation of the GRB emission should not be confused with the relativistic beaming of photons, which is due to the relativistic bulk motion of the emitting plasma (characterized by a Lorentz factor \( \Gamma \) which can be \( \sim 100 \) or more, see Chapters 5–6). Photons emitted isotropically in the rest-frame of a blob of radiating matter moving ultra-relativistically with a Lorentz factor \( \Gamma \), appear beamed in the observer’s frame within a cone of opening angle \( \theta_b \) along the direction of motion of the emitting blob. The relativistic beaming angle is \( \theta_b \sim 1/\Gamma \). The observer will see only those photons emitted by matter in the small cone defined by \( \theta_b \), and will have no way of determining the extension of the ejecta outside this narrow cone (see Figure 3.8). Consequently, there is no observable difference between the early afterglows of isotropic GRBs and those of beamed GRBs (as long as the beaming angle \( \theta_b \) is greater than \( \theta_c \)). Fortunately, the bulk Lorentz factor decreases with time and the reduction of the relativistic beaming allows us to see an increasing fraction of the radiating jet. When \( \theta_b \) becomes comparable with \( \theta_c \), the light-curves of the afterglows of beamed GRBs show an achromatic break that is absent in isotropic GRBs (Rhoads 1997).

There are various ways by which afterglows can be used to test the burst collimation. A first test is connected with the fact that \( \Gamma \) is lower when the late time afterglow is emitted than during the GRB. This implies that the jet has started to widen and that the emission is less beamed during the late afterglow than during the GRB itself. If GRBs are beamed, the lesser beaming of their afterglows will result in an afterglow rate exceeding the GRB rate. One consequence of this situation will be the existence of ‘orphan afterglows’, i.e. afterglows without GRB. We will come back to the possibilities to look for orphan afterglows at different wavelengths in the next paragraph. The detection of such events can help to measure the degree of collimation of GRB jets. Of course, these searches are complicated by the fact that these ‘orphan afterglows’ may be difficult to detect in the absence of the GRB signal. For instance, Mészáros, Rees, and Wijers (1998) have shown that there are many possibilities to have no optical afterglow: the density of the environment may be too low and the
external shock occur at larger radii and over much longer timescales, or the burst environment may be very dusty, in star-forming regions, and have a significant column density due to absorbing material.

Another test of the burst collimation was studied by Rhoads (1999a); it is based on differences between the dynamical evolution of collimated and spherical bursts. Burst ejecta decelerate through their interaction with the ambient medium. If the ejecta are initially emitted into a cone of opening angle $\theta_e$, the deceleration changes qualitatively when the bulk Lorentz factor of the beamed emission drops below $1/\theta_e$. Before this time, the Lorentz factor behaves as a power law in radius and the matter has not enough time (in its own rest-frame) to expand sideways. Afterward the Lorentz factor decays exponentially with radius. This change is due to the lateral expansion of the ejecta, which increases the rate at which additional material is accreted. Such lateral expansion is not possible in the spherical case. When the remnant reaches this exponential regime, when $\theta_b = \theta_e$, a break in the afterglow light-curve appears (Rhoads 1999a). Of course such breaks may not be uniquely attributed to jets (see the discussion in Frail, Waxman, & Kulkarni 2001). Moreover, such breaks are obtained considering a simple impulsive energy input characterized by a single energy and bulk Lorentz factor value.

Sari, Piran, and Halpern (1999) showed that the presence of a jet can be inferred, even in the absence of a break in the afterglow, simply by measuring the decay index of the afterglow (at the condition of knowing the electron energy distribution). Assuming a power-law electron spectrum with an index of $\sim$2.4, the authors expected, for high frequencies (above the synchrotron break frequency, $\nu_m$), a decay like $t^{-1.1}$ to $t^{-1.3}$ for spherical emission, and like $t^{-2.4}$ for a collimated emission (Table 1 of Sari, Piran, & Halpern 1999). They reported that GRB 980519 had the most
rapidly fading GRB afterglow observed at that time, decaying as \( t^{-2.05 \pm 0.04} \) in the optical (Halpern et al. 1999) and as \( t^{-2.07 \pm 0.11} \) in X-rays (Owens et al. 1998). This fast temporal decay was consistent with a spreading jet, and the authors suggested that the transition corresponding to 1/\( \Gamma \) \( \sim \theta_c \) took place during the few hours between the GRB and the first detection of the afterglow. This provides a strong argument for detecting GRB afterglows as soon as possible after the burst, because in the time window of few hours following a GRB the relativistic fireball can decelerate dramatically from \( \Gamma \sim 100 \) to \( \Gamma \sim 10 \). Thus, even without the detection of a clear break in the light-curve, the presence of a very rapid decline of the afterglow might be a strong indication for the presence of a collimated emission. Such a rapid decline, corresponding to a jet expanding laterally, could also explain the faint or absent optical afterglows of some GRBs, without invoking the presence of absorbing material. Swift would be well adapted to answer these crucial questions, thanks to the dramatic reduction of the delay between the GRB and the first observations of the afterglow at X-ray and optical wavelength (see Chapters 4 and 7).

GRB 990123 gives us another example of a possible jet. The optical afterglow from about 3.5h to \( \sim 2 \) days showed a power law decay with \( t^{-1.4 \pm 0.03} \). Then the optical emission began to decline faster (\( t^{-1.8} \)) (Fruchter et al. 1999, Kulkarni et al. 1999). This might indicate the transition to a sideways expanding jet. The transition took place at \( t_b \sim 2 \) days, corresponding to \( \theta_c \sim 0.1 \) according to the formula given below. This implies a beaming factor of about 200 (Sari & Piran 1999b). Like GRB 990123, the optical light-curve of the afterglow of GRB 990510 shows a break on a timescale of one day. Both bursts are among the highest fluence BeppoSAX events (Harrison et al. 1999). Quite remarkably the break is achromatic in the optical band (Figure 3.9, from Harrison et al. 1999), in agreement with the predictions of jet break models. The early decay of the radio afterglow, which occurs at \( t \sim 1 \) day, also favors the jet model. Kuulkers et al. (2000) and Pian et al. (2001) analyzed the X-ray afterglow and showed that its light-curve is also consistent with the jet interpretation. A double power law with indices \( \alpha_1 \sim 1 \) and \( \alpha_2 \sim 2 \) before and after a jet collimation break time of \( \sim 0.5 \) days after the GRB gives a good fit to the X-ray afterglow observation (Figure 3 of Pian et al. 2001). This was the first burst where there was evidence for a jet seen in the X-Ray afterglow light-curve. On the other hand, the optical afterglows of GRB 970508 and GRB 970228 showed no convincing evidence for breaks, even when observed after several months. They can be considered as un-collimated or only moderately collimated (Rhoads 1999a,b, Sari, Piran, & Halpern 1999), with \( \theta_c > 1 \) steradian (\( \theta_c > 57^\circ \)).

If the jets are connected with breaks and rapid decay slopes, the previous examples lead to values of \( \theta_c \) ranging from \( \theta_c < 0.1 \) for GRB 980519 to \( \theta_c \sim 0.1 \) for GRB 990123 and \( \theta_c > 1 \) for the two most famous bursts, GRB 970228 and GRB 970508.

A more complete analysis of a sample of 17 GRBs with known redshifts has been done by Frail et al. (2001) to search for beaming. In their paper they give the formula below, which derives the jet opening angle as a function of various parameters of the GRB (isotropic equivalent energy \( E_{iso} \), time of the jet break \( t_j \), redshift \( z \), efficiency to
convert the kinetic energy of the ejecta into gamma-rays $\eta_\gamma$, and of its environment (the density of the surrounding medium $n$):

$$\theta_j = 0.057 \left( \frac{t_j}{1 \text{ day}} \right)^{3/8} \left( \frac{1 + z}{2} \right)^{-3/8} \left[ \frac{E_{\text{iso}}(\gamma)}{10^{53} \text{ erg}} \right]^{-1/8} \left( \frac{n_\gamma}{0.2} \right)^{1/8} \left( \frac{n}{0.1 \text{ cm}^{-3}} \right)^{1/8} \quad (3.1)$$

The jet angles vary from $1^\circ$ to more than $25^\circ$, with a clear concentration near $4^\circ$ (Figure 3.10). The authors also note that, for this limited sample, very narrow
opening angles $\theta_j < 3^\circ$ are rarely required, for less than 10% of the BeppoSAX GRB sample.

Knowing the beaming angles of these GRBs and their redshift, and assuming uniform emission across the jet, it has been possible to determine $E_\gamma$, the true energy radiated in high-energy photons. $E_\gamma$ is calculated with the formula: $E_\gamma = f_b E_{iso}$, where $f_b = (1 - \cos \theta_j) \sim \theta_j^2/2$ is the beaming factor, and $E_{iso}$ is the isotropic equivalent energy. $E_{iso} = 4\pi D_L^2 S_\gamma (1+z)^{-1}$, where $S_\gamma$ is the fluence of the burst, $z$ its redshift and $D_L$ its luminosity distance. Frail et al. (2001) found the remarkable result that the values of $E_\gamma$ are clustered around $5 \times 10^{50}$ erg, with a much smaller dispersion than $E_{iso}$ (Figure 3.11). A very important point can be deduced from this beaming study: the central engines of GRBs release a similar amount of energy, with a significant part ($\sim 10^{51}$ erg) escaping as gamma rays (Frail et al. 2001). They explain the wide distribution of the observed fluences and $E_{iso}$ by a wide range of jet opening angles. If $\eta_\gamma$ is the efficiency of the fireball to convert the (kinetic) energy of the ejecta into gamma-rays, Frail et al. (2001) infer that, with $\eta_\gamma \sim 0.2$ and $E_\gamma \sim 5 \times 10^{50}$ erg, the true total energy of the fireball would be $\sim 3 \times 10^{51}$ erg, only slightly larger than the energy released by an ordinary supernova. From the observed GRB rate and the measure of the average beaming factor of GRBs, the observed GRB rate can be calculated as $R_{obs} = 0.5$ Gpc$^{-3}$ yr$^{-1}$ ($z = 0$), and the true rate as $R_t \sim 250$ Gpc$^{-3}$ yr$^{-1}$ ($z = 0$), with a mean value of $f_b^{-1} \sim 500$. The collapsar scenario discussed in Chapter 8 can easily provide a sufficient number of progenitors, but the coalescence of binary neutron stars is just capable of doing it. The estimated rate of NS coalescence is $R_c = 80$ Gpc$^{-3}$ yr$^{-1}$ (at $z = 0$), while the estimated rate of type Ibc supernovae is
Figure 3.11. Distribution of the apparent isotropic energy $E_{\text{iso}}$, for GRBs with known redshifts (top panel). The distribution of the beaming corrected energies is given in the bottom panel. The arrows indicate GRBs for which there are upper or lower limits only. The mean value of $E_{\gamma}$ is $\sim 5 \times 10^{50}$ erg and the narrowness of this distribution places restrictions on the dispersion of $n$ (the density of the surrounding medium) and $\eta$ (the gamma-ray efficiency). At the time of this study these two figures seemed to indicate that most of the dispersion in luminosity was due to the diversity in opening angles (from Frail et al. 2001).

much larger $R_{\text{lsr}} \sim 6 \times 10^4$ Gpc$^{-3}$ yr$^{-1}$ (Phinney 1991). The simple jet model used by Frail et al. (2001) is quite appealing but it is based on assumptions that are probably too simple to reflect the complexity of GRB emission. This complexity appears in the light-curves of many afterglows which show bumps and wiggles and which cannot be fitted with simple or broken power laws (see Lipkin et al. (2004) for an example). Obtaining the beaming angle of GRB jets might thus be more complicated. Nevertheless, this was a first attractive approach (see Chapters 4 and 7 for additional comments on the crucial question of the achromaticity of the breaks, and on the much more complicated situation revealed by Swift). Section 3.5 on orphan afterglows will indicate other ways to approach the question of GRB beaming.

3.5 ORPHAN AFTERGLOWS

The search for orphan afterglows was proposed as a means to measure the beaming of GRBs by J. Rhoads in 1997, and the first searches started soon afterwards, in 1998.

Orphan radio afterglows were searched for by Perna and Loeb (1998). From the total number of $>0.1$ mJy sources observed at 8.44 GHz (Windhorst et al. 1993) and the fraction of fading sources at 1.44 GHz (Oort & Windhorst 1985), they got a crude limit on the mean collimation angle, $\theta_c > 6^\circ$. Constraints on beaming were also looked for by Grindlay (1999), using the ‘fast transient’ X-ray sources from the Ariel 5 survey and the HEAO-1 experiments. He concluded that there is no strong beaming difference between the prompt GRB and the early X-ray afterglow. Nevertheless, a difference by a factor of 3 was still allowed by this study. A similar conclusion was reached by Greiner et al. (2000) after a search for transient X-ray sources in the Rosat All Sky Survey.
Rees (1999), using two studies to find supernovae down to $R \sim 23$ (Garnavich et al. 1998, Perlmutter et al. 1998) deduced a visible afterglow rate smaller than few times 0.1 sq. deg$^{-1}$yr$^{-1}$. The magnitude limit of these searches allowed the detection of optical counterparts to GRBs brighter than 1 ph cm$^{-2}$ s$^{-1}$. The GRB rate above this peak flux was measured by BATSE to be $\sim 300$ yr$^{-1}$, or 0.01 sq. deg$^{-1}$ yr$^{-1}$, and Rees concluded that the ratio of orphan afterglows to GRBs is unlikely to exceed $\sim 20$. The result is that $\Omega_\gamma > 0.05 \Omega_{\text{opt}}$ and, with $\Omega_{\text{opt}} > 0.4$, that $\Omega_\gamma > 0.02$ (about $5^\circ$). This crude estimation allows saying that the opening angle of the ultra-relativistic $\gamma$-ray emitting material is no less than $\sim 5^\circ$ (Rees 1999). This is in relatively good agreement with the mean value found by Frail et al. (2001; see Section 3.4).

More recent studies are described in Chapter 9. These studies have all been negative, but they have provided stronger constraints on the overall rate of orphan afterglows. Such orphan searches are of prime importance because the collimation of the jets has a strong impact on the energy budget of the progenitor (reducing it by the collimation factor $f_b \sim \theta_b^2/2$, $f_b \sim 5 \times 10^{-3}$ for $\theta_b \sim 0.1$ rd), and on the GRB burst rate (which increases by the inverse factor $1/f_b$). A review of orphan afterglow searches and perspectives is given in Rhoads (2004).

### 3.6 THE GRB–SUPERNova CONNECTION

Considering the energy of GRBs ($10^{51}–10^{52}$ erg), it is difficult not to think of supernova explosions, which have a comparable energy budget, with an energy dissipation occurring on a much longer timescale. As we have seen (see also Chapter 5), the baryon mass involved in a GRB explosion cannot be large ($10^{-3} M_\odot$ at most, and probably much less), in order to avoid the thermalization of the energy, and to preserve the possibility of having large Lorentz factors ($\Gamma > 10^2$). The possible connection between the two phenomena was considered by Woosley (1993), and by McFadyen and Woosley (1999), invoking collapsars—‘failed’ supernovae with a massive core—as the energy source of GRBs.

The first observational evidence of a possible connection between SN and GRB was discovered by Galama et al. (1998a), when they noticed that the WFC error box of GRB 980425 contained a new supernova, SN 1998bw. This SN exploded in the nearby face-on barred spiral galaxy ESO 184-G82, at a distance of $\sim 40$ Mpc ($z = 0.0085$). Galama et al. showed that the probability of catching a supernova purely by chance, in any of the 13 GRB error boxes obtained by the WFC at that time, was quite low, of the order of $9 \times 10^{-5}$ (Galama et al. 1999b). In the WFC error box of GRB 980425 there were two weak X-ray sources: the first, called SAXJ 1935.0-5248, was situated 50″ from SN 1998bw and consistent with it; the second, called 1SAXJ 1935-3-5252, was 3′ off SN 1998bw and inconsistent with it (Piro et al. 1998b, Pian et al. 1999, Galama et al. 1998a). The time of SN collapse, derived from the accurate modeling of its light-curve, coincided with that of the GRB to within about 1 day ($+0.7$, $-2.0$ days). Due to its very small distance, GRB 980425 was a long GRB with an energy release of only $E_{\text{iso}} = 9 \times 10^{47}$ erg. This was much smaller, by four orders of magnitude, than ‘standard’ cosmological GRBs. This energy is close to the
energy, calculated by Colgate in 1974, that is released when a supernova shock wave breaks through the surface of a large radius star. But SN 1998bw was profoundly different from supernovae known by then: it was of type Ic,¹ and not of type II as proposed by Colgate, and its peak luminosity was unusually high compared with typical type Ic supernovae. SN 1998bw was also remarkable for its very high radio luminosity; at early times it was the brightest SN ever observed at these wavelengths (Kulkarni et al. 1998b). It was equally remarkable for the energy of its sub-relativistic ejecta: $(2-6) \times 10^{52}$ erg, a factor of $\sim 30$ higher than the energy of a typical type Ib/c supernova, and by its large production of $^{56}$Ni: $0.5-0.7 M_\odot$, which had no precedent for a core collapse supernova (Woosley, Eastman, & Schmidt 1999). This supernova was thus called a ‘hypernova’. The extremely large energy involved in the explosion may suggest the existence of a new mechanism of massive star explosion that could produce the relativistic shocks necessary to generate the observed GRB.

Iwamoto (1999) found that the SN light-curve could be well reproduced by an extremely energetic explosion of a massive star with initial mass of $40 M_\odot$, composed mainly of C and O (12–15 $M_\odot$). Woosley, Eastman, and Schmidt (1999) considered the explosion of a 6 $M_\odot$ core of C and O. The mass of the core after the explosion was estimated to be larger than 3 $M_\odot$, and therefore the remnant had to be a black hole. A major question was: How could this SN generate a GRB? The models which explained the optical observations of SN 1998bw (Iwamoto et al. 1998, Woosley, Eastman, & Schmidt 1999) were unable to explain the GRB, since the ejecta of supernovae are characterized by non-relativistic velocities. Kulkarni et al. (1998b), using radio data, concluded that SN 1998bw had a shock wave moving at relativistic speed ahead of the low-velocity ejecta that would power the optical light-curve. This relativistic shock could potentially generate a GRB at early times when it collides with the circumstellar or interstellar matter. The acceleration of non-thermal electrons at the shock front would result in the emission of high-energy photons via synchrotron emission.

Other observations of BeppoSAX GRBs supported this link between SN and GRBs. Bloom et al. (1999b) discovered a late optical bump superimposed on the power law decay of the optical afterglow of GRB 980326, which could be due to an underlying supernova. In fact, they showed that the observed optical light-curve

¹ Supernovae are classified into types according to the elements present in their optical spectrum: the spectra of Type II supernovae show hydrogen Balmer lines in absorption and/or emission at maximum brightness, while the spectra of Type I supernovae do not show hydrogen Balmer lines. Type I supernovae are further subdivided into Types Ia, Ib, Ic. Supernovae that show a strong ionized silicon line at 6150 angstroms are classified Ia. Type I supernovae lacking this Si II feature are classified as Type Ib if they show substantial evidence of neutral Helium, especially the neutral helium line at 5876 Angstroms. If the neutral helium features are weak or missing entirely then the supernova is classified as Type Ic. Type II and Ibc supernovae are core-collapse supernovae, resulting from the collapse of a massive star. Type Ia supernovae are the consequence of the accretion-induced collapse of a white dwarf in a binary system.
could be the combination of a power law (with an exponent equal to \(-2\)) and a bright SN at a redshift \(z \sim 1\). Thirty days after the GRB, the optical flux was a factor of 10 above the extrapolated power law decline (Figure 3.12). This increase could not be due to the host galaxy, because it was not present at later times (Bloom & Kulkarni 1998). A deviation of the optical afterglow of GRB 970228, the first BeppoSAX burst with an afterglow, from a power-law decay, was also interpreted in a similar manner by Reichart (1999) and Galama et al. (2000) (Figure 3.4b). Other events might be associated with supernovae, among them: GRB 980703 (Holland et al. 2001), and GRB 990712 (Fruchter et al 2000a, Björnsson et al. 2001). GRB 011121 provides
Figure 3.12. (Opposite page) R-band light-curve of the afterglow of GRB 980326, which can be fitted with a bright supernova superimposed on a power law afterglow. Since the redshift of the GRB is not known, the authors consider different redshifts for the supernova (Bloom et al. 1999b). The light-curve of the well observed supernova SN 1998bw (Galama et al. 1998a) was used as a template for the SN contribution. The GRB + SN model at a redshift of about unity provides an adequate fit to the data. (This page) Another example of a possible SN contribution in the light-curve of GRB 011121. The dotted lines show the optical afterglow power-law decay. The dashed line shows the R-band contribution of a supernova like SN 1998bw, redshifted to $z = 0.36$, corrected for extinction, and scaled down by 0.1 mag (Garnavich et al. 2003). The three isolated points at 8–10 days are J-band observations. The points after 10 days are R-band observations. The solid line is the sum of the afterglow and the possible supernova contribution.

another case for a late optical bump in the light-curve of the afterglow (Bloom et al. 2002; see their Figure 2). This burst has been considered by the authors as the best case for an underlying supernova, giving compelling evidence for a massive star origin of long-duration GRBs. Moreover, they concluded that models invoking a supernova explosion occurring weeks to months before the GRB (the supranova model; Vietri & Stella 1998) were excluded by these observations. Finally, if the bump was due to a SN, it could not be explained by a redshifted SN 1998bw contribution, the SN associated with GRB 011121 had to be bluer, slightly fainter and faster than SN 1998bw (Garnavich et al. 2003; Figure 3.12).
Motivated by the association of GRB 980425 with SN 1998bw, Bloom et al. (1998b) discussed the possible existence of a sub-class of GRBs produced by supernovae. Kippen et al. (1998) investigated the possibility that a fraction of BATSE GRBs could be associated with supernovae, using the BATSE and BATSE/Ulysses locations. They found that only a very small fraction (less than 0.2%) of high-fluence BATSE bursts could come from known supernovae. Kommers et al. (2000), using the fact that nearby supernovae have a homogeneous distribution in a Euclidean space \( \langle V / V_{\text{max}} \rangle = 0.5 \), estimated that the SN component in GRBs can contribute at most 10% of the observed BATSE GRB sample. Graziani, Lamb, and Marion (1999), using a similar argument, estimated that no more than 5% of the GRBs detected by BATSE were produced by SN Ib/c. This is perhaps not surprising because GRBs with isotropic equivalent energies of \( \sim 10^{48} \text{erg} \) (as GRB 980425), associated with supernovae, would remain undetectable with current instruments, even if they were quite common. In fact, this last study was done at a time where the association of GRB 980425 with SN 1998bw was largely debated, and these authors tried to demonstrate that considerable caution was needed before claiming such associations. Later, Zeh, Kloke, and Hartmann (2004) tackled this question with a different approach: they searched SN light in the optical light-curves of 21 nearby GRB afterglows. They found that the optical afterglows of all GRBs with redshift smaller than \( z \sim 0.7 \) showed a weak excess at late time, supporting the possibility that all afterglows of long-duration GRBs contain light from an associated supernova. They found no correlation of the supernova luminosity with the properties of the GRB and its optical afterglow.

Later, the detection of supernovae 2003dh and 2006aj, respectively associated with GRB 030329 (HETE-2) and GRB 060218 (Swift) would revive this question (see Chapter 4). The best constraints today on the fraction of supernovae producing GRBs have been obtained with radio observations. Soderberg et al. (2006), for instance, found that less than 10% of type Ibc SNe are associated with GRBs.

Even if it is difficult to measure the fraction of supernovae producing GRBs, it is certain that not every supernova can make a strong burst, not even those which are powered by black-hole accretion (Woosley, Eastman, & Schmidt 1999). Connected with this question, it is worth noting that there is about one SN explosion every second in the observable Universe, while GRBs are much more rare, with one or two events observed each day (and at most \( 10^3 \) GRBs produced per day if the beaming factor is \( \sim 500 \)). The models proposed by Woosley (1993) and Paczyński (1998), which assume a black hole surrounded by a massive torus, support such associations with SNe. But many key questions remain, concerning the rare circumstances in which SNe are able to produce GRBs, the role played by rotation and collimated jets in the generation of relativistic flows with high Lorentz factor etc. More observations of nearby GRB–supernova associations would be available in the HETE/Swift period, allowing us to come back to this question in Chapters 4 and 8. This is an issue of fundamental importance for the identification of GRB progenitors since the supernova signature is a clear discriminator between merger and massive star scenarios.
3.7 EMISSION AND ABSORPTION FEATURES IN THE X-RAY SPECTRA OF GRBS AND IN THEIR X-RAY AFTERGLOWS

X-ray lines have always provided powerful tools for the study of the physical and chemical state of hot plasmas (in active galactic nuclei, galactic X-ray binaries, warm intergalactic gas confined in the potential well of clusters of galaxies, etc.). The detection of X-ray lines in the prompt GRB emission or in their early afterglows would provide a fantastic diagnostic of the close environment of these sources and of their possible progenitors. As the iron line is the most common feature in X-ray sources, most of the searches have been concentrated on this element. They include searches for the K$_\alpha$ lines from neutral to H-like Fe ions, with rest-frame energies of 6.4 to 6.9 keV, searches for the recombination edge in emission from H-like ions at 9.3 keV (rest-frame), and searches for the absorption edge from neutral iron in absorption at 7.1 keV (rest-frame). The following detections have been reported by the BeppoSAX team:

(1) BeppoSAX WFC and GRBM detected on July 5th, 1999 a GRB which was the second brightest in $\gamma$-rays. It lasted 42 s in $\gamma$-rays and ~60 s in X-rays, and presented a typical hard-to-soft evolution. An absorption feature at 3.8 ± 0.3 keV was visible in the second time interval of 7 s. Amati et al. (2000) attributed this absorption feature to the photo-ionization K-edge of neutral iron. They proposed that neutral iron could be first photo-ionized by the burst of X-ray photons until the electrons are stripped from the iron ions, causing the disappearance of the feature. The observed edge at 3.8 ± 0.3 keV would be due to the neutral iron edge (7.1 keV) for a source at a redshift: $z = 0.86 ± 0.17$. Optical observations of the host galaxy performed later measured a redshift compatible with the value inferred from X-rays: $z_{\text{opt}} = 0.8424$ (Le Floc’h et al. 2002). One potential problem of this explanation is that it requires huge iron abundance (~75 times the solar one). This iron abundance points to the existence of a burst environment highly enriched by a SN (Amati et al. 2000). This favors models in which GRBs originate from the collapse of very massive stars and are preceded by a SN-like explosion. The explosion of the GRB within young supernova remnants (Lazzati et al. 2001) is compatible with the supranova scenario of Vietri and Stella (1998), a variation of the hypernova scenario, where the GRB follows a SN explosion (by weeks to years), with the ejection of an iron-rich massive shell (see Chapter 8).

(2) Four additional bursts have shown evidence for an iron emission line during the X-ray afterglow observed 8–40 h after the burst event. These bursts are GRB 970508 (Piro et al. 1999), GRB 970828 (Yoshida et al. 1999), GRB 991216 (Piro et al. 2000), and GRB 000214 (Antonelli et al. 2000). An emission line feature was observed in GRB 000214 with a centroid energy of 4.7 ± 0.2 keV. If this emission feature is interpreted as Fe K$_\alpha$, the corresponding redshift is between $z = 0.37$ and 0.47, but no optical or near-infrared afterglow was identified during follow-up observations (Rhoads et al. 2000), preventing the measure of the redshift of this burst. This lack of counterparts may result from extinction in a dense
surrounding medium. For GRB 970508, a possible Fe Kα line is observed at 3.4 ± 0.3 keV. In this case the redshift of the host galaxy was measured: \( z = 0.835 \) (Metzger et al. 1997b). Applying this redshift to the line, its corresponding energy in the source rest-frame is 6.2 ± 0.6 keV. The line disappeared just after a re-burst in the X-ray (2–10 keV) light-curve of the afterglow. This was not the case for GRB 000214 where no re-burst was observed.

The observation of these features is important because they imply a rich environment, very close to the GRB region. Of course they favor the hypernova or supranova scenarios. The iron-rich environment may be produced by the stellar wind before the explosion of the hypernova (Mészáros & Rees 1998), or deposited after a SN explosion occurring few months before the burst in the supranova scenario (Vietri & Stella 1998). But for both models, and in the case of these two GRBs, the problem is to deposit quite a large amount of iron at a large distance from the source, typically \( \sim 3 \times 10^{15} \) cm. The densest circumstellar envelopes around the progenitors of some peculiar core collapse supernovae such as SN 1997ab imply masses at such distance of the order of \( M \sim 10^{-1} \) to \( 10^{-2} \) \( M_\odot \), largely below the mass of about one solar mass needed in the emitting region of GRB 000214 (Antonelli et al. 2000).

Using the large X-ray space observatories XMM-Newton and Chandra, the afterglows can be studied with a much better sensitivity and spectral resolution. For instance Reeves et al. (2002) reported the detection of emission lines of magnesium, silicon, sulphur, argon, and calcium from metal-enriched material in the X-ray spectrum of the afterglow of GRB 011211. The XMM observation started 11h after the initial burst. The spectroscopy of the optical afterglow allowed measuring the redshift \( z = 2.141 \pm 0.001 \) (Fruchter et al. 2001, Holland et al. 2002). The probable host galaxy has a magnitude \( m_v = 25 \pm 0.5 \). Figure 2 of Reeves et al. (2002) shows the EPIC spectrum of the burst afterglow with the reported emission lines (Figure 3.13). In this event, the absence of Fe K emission can be explain by a short delay (\( \sim 4 \) days) between the SN explosion, which has ejected the outflowing matter, and the GRB. This result contrasts with the large mass of iron implied by the previously reported iron lines in the X-ray spectra of other GRBs, which have been explained by a much longer delay between an initial supernova and the onset of the GRB several months later. These observations provoked an intense debate in the community concerning both the reality of the features detected (Borozdin & Trudolyubov 2003, Rutledge & Sako 2003, the response of Reeves et al. 2003, Butler et al. 2005), and the interpretation of the X-ray spectrum (Tavecchio, Ghisellini, & Lazzati 2004). It was expected that the launch of Swift with an X-ray telescope arriving on the burst a few minutes after the trigger (instead of hours for BeppoSAX or XMM) would solve the question of the existence of such lines. In this respect, we believe that it is fair to say that no convincing lines were reported in the X-ray afterglows of Swift GRBs detected during the first four years of the mission.
Figure 3.13. The X-ray spectrum of the afterglow of GRB 011211 at redshift $z = 2.14$, measured with XMM-Newton. Significant line emission was observed during the initial 10 ks and the [0.2–3] keV spectrum shown above is obtained for the first 5 ks of exposure (Reeves et al. 2002). Emission lines are detected at the following energies (in kiloelectronvolts) in the burst rest frame: $1.40 \pm 0.05$, $2.19 \pm 0.04$, $2.81 \pm 0.04$, $3.79 \pm 0.07$, and $4.51 \pm 0.12$. The frequently observed K$_\alpha$ transitions which are close to the observed lines are: Mg XI (1.35 keV) or Mg XII (1.47 keV), Si XIV (2.00 keV), S XVI (2.62 keV), Ar XVIII (3.32 keV), and Ca XX (4.10 keV). The best-fit redshift for the line set was found to be $z = 1.88 \pm 0.06$, differing from the GRB redshift $z = 2.14$. The lines are thus blueshifted with respect to the host of GRB 011211. This implies an outflow velocity $v \sim 25 \, 800 \, \text{km} \, \text{s}^{-1}$ of the line-emitting material. The authors considered that the line emission was detected with good confidence (Reeves et al. 2002).

3.8 X-RAY-RICH GAMMA-RAY BURSTS

Another interesting result provided by BeppoSAX was the identification of a new sub-class of GRBs. They were called the fast X-ray transients or X-ray flashes (Heise et al. 2001). They are flashes of X-rays, so far not recurrent (different from the soft gamma-ray repeaters), sometimes accompanied by weak gamma-ray emission with no properties common to any known class of X-ray sources. The unique property of these bursts is their large X-ray flux comparable to or even dominant over their gamma-ray emission. They were defined by Heise et al. (2001) as ‘bright X-ray sources lasting of the order of minutes, which are detected in the Wide Field Cameras in the energy range 2–25 keV, but remain undetected in the Gamma-Ray Bursts Monitor on BeppoSAX’. Such events had been previously reported by Strohmayer et al. (1998), using Ginga data (see Chapter 2). They found for some events, a large X-ray fluence, much larger than the few percent usually measured in classical GRBs.
X-ray flashes (XRF) seemed to represent approximately one-third of the GRBs detected with the WFC (Feroci et al. 2001). In their discovery paper Heise et al. (2001) discussed the possible origin of these X-ray flashes: they rejected the hypothesis of highly redshifted GRBs (which could explain the low value of $E_{\text{peak}}$) because the X-ray flashes in their sample did not appear time-dilated. They favored the hypothesis of events with lower Lorentz factors than average, which they attributed to a larger baryon loading than for classical GRBs (e.g. Dermer, Chiang, & Böttcher 1999).

Along with the discovery of X-ray flashes, which were seen in the WFC but not in the GRB Monitor, BeppoSAX reported the detection of few events with an unusually high proportion of X-rays. Two examples of such GRBs are GRB 981226 (Frontera et al. 2000) and GRB 990704 (Feroci et al. 2001). For this last burst the peak flux ratio between the X-ray and the $\gamma$-ray domains (2–10 keV/40–700 keV) was $F_x/F_\gamma \sim 0.6$, and the fluence ratio in the same energy bands was $S_x/S_\gamma \sim 1.5$. The flux and fluence ratios for GRB 981226 were a little lower, but both values are well above the values of classical GRBs (see Figure 3 of Feroci et al. 2001). Another interesting case is GRB 990712 (Frontera et al. 2001). This event is a clear mixture of a classical GRB with an X-ray flash. The first pulse (duration <10 s) is very hard with a peak energy above the energy pass band of BeppoSAX ($E_{\text{peak}} \geq 700$ keV). The second pulse, between 10 s and 20 s, is quite soft with $E_p \sim 10$ keV. The $\gamma$-ray (40–700 keV) fluence of the burst is $S_\gamma = (6.5 \pm 0.3) \times 10^{-6}$ erg cm$^{-2}$, and its X-ray (2–10 keV) fluence is $S_x = (2.6 \pm 0.06) \times 10^{-6}$ erg cm$^{-2}$ with a ratio $S_x/S_\gamma = 0.40 \pm 0.03$. This GRB has one of the lowest redshifts ($z = 0.43$) of BeppoSAX bursts (GRB 980425 apart, see Section 3.6), and it is also one of the least energetic events. Such events might indicate continuity between typical GRBs and XRFs as a new class of very soft GRBs (Heise et al. 2001). This was also the conclusion of Kippen et al. (2001, 2003, 2004) who rediscovered nine of BeppoSAX X-ray flashes in BATSE continuous data. They found that the durations and spectra of these events were consistent with the extrapolation of the BATSE population towards faint and soft events, suggesting that X-ray flashes constitute a natural extension of the previously known GRB population.

Finding the causes of the softness of XRFs is of great importance in understanding the basic parameters which determine $E_{\text{peak}}$, the peak energy of GRBs (see Section 3.10). In this respect important efforts were made to study the other properties of X-ray flashes, like their afterglows or host galaxies (Frail et al. 1999, Bloom et al. 2003). For GRB 990704 the X-ray afterglow is particular, with a very slow power-law decay (index $\sim 0.8$), the average value of the power law being 1.36 for the 13 GRB X-ray afterglows observed between 1997 and 1999 (Feroci et al. 2001). For GRB 981226, its X-ray afterglow was also peculiar with an initial short rise or may be a plateau never seen in any other afterglow (Frontera et al. 2000). For these two bursts no optical afterglow was reported, whereas a radio afterglow was observed for the last one (Frail et al. 1999). Despite some differences, the basic properties of XRFs were found to be similar to those of classical ‘hard’ GRBs. This will also be the conclusion derived from the study of HETE-2 XRFs discussed in the next chapter, along with the proposed theoretical explanations of X-ray flashes.
3.9 THE CONNECTION BETWEEN THE PROMPT GRB AND THE AFTERGLOW

Of the GRBs promptly localized with BeppoSAX, 90% display a detectable X-ray afterglow, less than half also exhibit optical emission, and only about 20% show radio emission. The fireball model explains the gross characteristics of the afterglows, like the power-law decay of the X-ray and optical emissions or the delay needed by radio afterglows to reach their peak.

In X-rays, the prompt and afterglow emissions really seem to be connected, and the question is to know when the afterglow starts. The various possibilities can be illustrated by few examples.

For GRB 970228 (see Costa et al. 1997e and Section 3.2) the train of three pulses following the first main peak has a flux consistent with the backward extrapolation of the afterglow, and Soffitta et al. (2004) considered that these pulses are already part of the afterglow. The temporal gap between the GRB (first peak) and the afterglow (the three pulses) is quite short, but real if this interpretation is correct. Such a short delay was also reported for GRB 980613 with a temporal gap of 30 s to 50 s between the GRB and the afterglow.

In other cases there is continuity in the temporal evolution from the main event to the NFI power-law light-curve. This is observed for GRB 970508 (Piro et al. 1998a) and for GRB 980923. In gamma-rays, this last burst showed a rapid variability for $\sim 40$ s followed by a power-law emission tail which lasted $\sim 400$ s. The spectro-temporal properties of this tail are quite different from the rapidly varying phase but very similar to those of the afterglow at lower energy (Giblin et al. 1999). This is consistent with the theoretical predictions of Sari and Piran (1999a). If the X-ray emission is attributed to the external shock of the jet on the surrounding medium, it means that this external shock must appear very early, even during the prompt $\gamma$-ray burst phase.

The transition between the prompt emission and the early afterglow has been also observed in GRB 011121 and GRB 011211 (Piro et al. 2005). For these two bursts a late X-ray flare takes place hundreds of seconds after the prompt emission. Its spectrum is softer than the prompt emission and quite similar to the power law observed in the afterglow at later times. Moreover, the tail of the late X-ray flare is connected with a single power law to the afterglow at 1 day. For these two events, the authors concluded that the late X-ray flares represent the beginning of the afterglow. Another example, which shows the detection of the afterglow in gamma-rays is given by Burenin et al. (1999) for GRB 920723. This burst is similar to GRB 980923, with a long soft gamma-ray tail lasting for at least 1000 s after a short main pulse (6 s long). This tail is characterized by a spectrum softer than the first pulse and a power-law decay $F \propto t^{-0.7}$, much flatter than classical afterglow tails measured with BeppoSAX. In this burst, too, there is a continuous transition between the GRB and its afterglow, observed at high energies. In addition, there is a clear connection between the tail of the first pulse (6 s after the trigger) and the power-law decay measured at late times during the gamma-ray tail (20–1000 s after the trigger (Figure 3.14).
Figure 3.14. The light-curve of GRB 920723 observed with Sigma on GRANAT in the 35–300 keV energy range. The reference time is the time of GRB trigger. This burst presents a continuous transition between the main burst and its power-law gamma-ray afterglow, indicating that the GRB enters the afterglow phase very soon after the prompt gamma-ray event (Burenin et al. 1999).

There are also some GRBs showing a possible X-ray break, the last portion of the GRB light-curve decaying faster than the backward extrapolation of the NFI afterglow light-curve (Soffita et al. 2004). In this case a broken power law is a better fit between the WFC points and the NFI ones. This is the case of GRB 990510 (Plan et al. 2001), where the WFC light-curve decays faster than the afterglow (see Costa and the BeppoSAX GRB Team 1999 for a review of the properties of 14 X-ray afterglows observed with the BeppoSAX WFC and NFI during the 1997–1998 period).

Despite these observations, BeppoSAX was not ideal to study the transition between the prompt emission and the afterglow, because it needed a re-pointing to place the GRB within the field of view of the narrow field instruments. Usually, the GRB disappeared below the detection threshold of the WFC after a few minutes and it could not be recovered in the field of view of narrow-field instruments for many hours. Even if BeppoSAX was 10 to 20 times faster than the other spacecraft in operation at that time, it took a minimum of 5 to 8 h to perform the complete chain of operations leading to the observation with narrow-field instruments: detect the GRB,
get the data to the ground, localize the burst, send the command to re-point the spacecraft in the direction of the X-ray afterglow (see Section 3.1). This gap would be filled after 2004 with the launch of Swift (see Chapter 4), which would autonomously re-point its narrow-field instruments to the source in less than 2 min. This will disclose a completely new, and mostly unexpected, phenomenology which illustrates the fact that new windows on GRBs always reveal surprises.

Regarding the optical afterglows, the observations started in general after the WFC error-box determination, hours after the GRB. One remarkable exception was GRB 990123 for which simultaneous observations during the GRB revealed a bright optical emission (9th magnitude). The story of this detection illustrates both the specific difficulties associated with GRB observations and the required luck for significant discoveries. GRB 990123 was detected and localized quickly by BATSE with a precision of 10°. Just 22 seconds after the BATSE trigger, the ROTSE robotic instrument (Akerlof et al. 1999) pointed its 16° × 16° field of view towards the error box and started taking pictures of the region. By chance this burst also triggered BeppoSAX which provided a much more accurate localization (5 arcmin) when the data reached the ground, 4 hours after the burst. This refined position was 11° off the centre of the BATSE localization and near the edge of the ROTSE images. The examination of ROTSE images, recorded in the seconds following the trigger, revealed a bright varying optical source at the exact position of the optical and X-ray afterglows, which had been found in the meantime. This was the first observation of the optical emission of a GRB during the prompt phase. This emission was extremely bright, reaching 9th magnitude for a source at redshift \( z = 1.6 \) (Figure 3.15). This observation will remain unique for many years because BeppoSAX was not able to feed the robotic instruments with fast, precise localizations. As we have said, the localizations of the WFC, which were measured on the ground, arrived few hours after the GRB.

The connection between the GRB itself and its afterglow would be best studied with the successors to BeppoSAX. After 2000, HETE-2 would feed the robotic telescopes with arcminute positions a few seconds after the burst, allowing the study of the transition between the GRB and its afterglow in the optical domain. After 2004, Swift would continue to quickly alert the robotic telescopes and get simultaneous optical and X-ray observations starting about 1 minute after the burst.

### 3.10 THE \( E_{\text{peak}} - E_{\text{iso}} \) (AMATI) RELATION

After the measurement of a dozen GRB redshifts, Amati et al. (2002) presented a study of the intrinsic spectral properties and energetics of this sample. Given their large distances, GRBs undergo significant cosmological effects which modify their observed properties. The intrinsic properties, obtained after correction for the cosmological effects, can be used for important studies like the inter-comparison of GRBs, or the study of GRB evolution with redshift. The parameters studied by Amati et al. (2002) included the duration and the spectral parameters of the prompt emission, described by two smoothly connected power laws (the so-called
Figure 3.15a. The optical light-curve of GRB 990123. The very early points of ROTSE I are indicated. The first analyzable image started 22 s after the GRB onset. The flux rose in the following 25 s, then a rapid decline began (with a power-law slope of $-2.5$), which gradually slowed to give a slope of $-1.5$ just 10 minutes after the burst. This decline agrees well with the power-law slope found when standard afterglow measurements were available, 3 h later (Bloom et al. 1999a). The inset shows the first three ROTSE optical measures compared to the BATSE $\gamma$-ray light-curve. ROTSE observations, which started 22 s after the onset of the burst, show an increase in brightness by a factor 14 during the first 25 s and decreases by a factor of 100 below the detection threshold $\sim 700$ s after the burst onset (Akerlof et al. 1999).

‘Band function’ with four parameters: $\alpha$, the low-energy photon index, $\beta$, the high-energy photon index, $E_{\text{peak}}$, the peak energy, and a normalization, see Band et al. 1993 and Chapter 2). The measurement of the spectrum allowed the determination of $E_{\text{iso}}$, the total (bolometric) energy radiated by the GRB during its prompt phase (Amati et al. called this quantity $E_{\text{rad}}$ but the usual notation is now $E_{\text{iso}}$). The authors performed a correlation study of the spectral parameters with the redshift $z$ and with $E_{\text{iso}}$. They found evidence for correlations between the following quantities: $\alpha$ and $z$, $E_{\text{peak}}$ and $z$, and $E_{\text{peak}}$ and $E_{\text{iso}}$, this last correlation being the most convincing (Figure 3.16). This correlation was to receive considerable attention in the years to come due to its potential for the determination of $E_{\text{iso}}$ independently of the redshift. Assuming a
Figure 3.15b. The R-band light-curve of the afterglow of GRB 990123 extended to 10 days. The dashed line indicates a power-law fit to the light-curve for $t > 1.1$ days, the exponent is $-1.12 \pm 0.03$. The last three ROTSE data points are in good agreement with the power-law fit of the afterglow (Galama et al. 1999a).

The $E_{\text{peak}} - E_{\text{iso}}$ (Amati) relation has been known for a long time and discussed by many authors. This correlation was observed in BATSE bursts (Mallozzi et al. 1995, Brainerd 1997), as well as in IPN bursts (Dezalay et al. 1997). In the absence of a distance determination it was difficult...
to assess the origin of this correlation (intrinsic or cosmological). The measurement of the first GRB redshifts provided sufficient information on their distance distribution to draw a conclusion about the intrinsic nature of the hardness–brightness correlation (Lloyd, Petrosian, & Mallozzi 2000, Atteia 2000).

The work of Amati et al. (2002) made possible the identification of the parameters involved in the correlation, the accurate determination of its slope, and the discovery of the small dispersion of GRBs around the best fit. The great progress resulting from this study was the discovery of the quality of the correlation. While it was already clear at that time that the burst luminosity was the key factor driving many other parameters, the correlation measured by Amati et al. was significantly tighter than previously known correlations, like the lag–luminosity relation discovered by Norris, Marani, and Bonnell (2000) or the variability–luminosity relation discovered by Reichart et al. (2001). The discovery of this relation triggered new fields of investigation concerning the searches for still tighter correlations, the construction of redshift indicators and the use of GRBs for cosmology.

In 2004, Yonetoku et al. (2004) discovered that $E_{\text{peak}}$ appears more tightly correlated with $L_{\text{iso}}$ than with $E_{\text{iso}}$. They used this relation to infer the redshifts of 680 GRBs in the BATSE catalog, in order to compare the history of GRB formation with the star formation history. The same year, Ghirlanda et al. (2004) plotted $E_{\text{peak}}$ vs $E_{\text{rad}}$, where $E_{\text{rad}}$ is the energy radiated in gamma-rays after correction for the beaming angle. They were inspired by the fact that $E_{\text{rad}}$ has a smaller dispersion than $E_{\text{iso}}$. They discovered that the $E_{\text{peak}}$–$E_{\text{rad}}$ relation is also more clustered than the Amati relation. This relation is, however, more difficult to establish because the determination of $E_{\text{rad}}$ requires the measurement of the beaming angle of the GRB which in turn relies on the detection of a clear achromatic break in the light-curve of the afterglow (see Section 3.4). Progress is continuing in this field with
the establishment of new tight correlations between intrinsic GRB parameters, for instance those discovered by Liang and Zhang (2005) and by Firmani et al. (2006) (see Chapter 9).

The parameters in these correlations are usually not affected in the same way by cosmological effects, opening up the possibility of inferring the redshift of a GRB, by calculating the redshift that brings it into the correlation. This possibility has been used to compute ‘pseudo-redshifts’ which can be used for statistical studies (e.g. Lloyd-Ronning, Fryer, & Ramirez-Ruiz 2002, Schaefer 2003, Band, Norris, & Bonnell 2004, Yonetoku et al. 2004). Pseudo-redshifts based only on the prompt emission offer the interesting potential to quickly identify nearby or very distant GRBs, which may deserve special follow-up (Atteia 2003, Pêlangeon & the HETE-2 Science Team 2006).

Some of these correlations are so tight that they may permit the measurement of the intrinsic luminosity of a GRB without knowing its redshift. If this can be done for a sufficient number of GRBs, it is possible to plot the Hubble diagram of GRBs, displaying GRBs in a luminosity–redshift diagram. This was suggested by Ghirlanda et al. (2004), using the $E_{\text{peak}}-E_{\text{rad}}$ relation, but other tight correlations could be used (see Liang & Zhang 2005, Dai, Liang, & Xu 2005, Friedman & Bloom 2005, and Firmani et al. 2006b). An excellent review of the potential of this field has been given by Ghirlanda, Ghisellini, and Firmani (2006). In the last chapter of this book (Chapter 9), we will come back to the crucial problem of the intrinsic luminosity of GRBs and to the various correlations which have just been evoked here. Improving and understanding them is a prerequisite to using GRBs as cosmic rulers, but for the time being, it is reasonable to say that, contrary to SNIa, GRBs cannot be considered to be standard candles.

3.11 CONCLUSION

In this chapter, we have seen some major discoveries of BeppoSAX. All of them concern long GRBs, because BeppoSAX was not able to localize the short GRBs with the WFC. Among them, the major result is, beyond doubt, the discovery of the afterglows in the X-ray band, which enabled identification of optical and radio afterglows. The precise determination of the positions of optical afterglows made it possible to determine the redshifts of the sources and identify their host galaxies, thanks to quick, efficient follow-up campaigns with very large telescopes. For the first time, and after 20 years of unsuccessful counterpart searches, the cosmological distances of GRBs could be firmly established. This was really a major step forward in GRB studies. Very early (1997) a GRB detected at a redshift $z \sim 3.4$ indicated that GRB explosions were already present in the young Universe, only 2 billion years old in this case. Another result of BeppoSAX was the absence of optical afterglows for nearly 50% of the GRBs, leading to the notion of dark bursts. We have discussed how they can be explained.

Following the first measurements of GRB redshifts, it appeared that the amount of energy associated with some of these GRBs was larger than the rest mass energy of
a solar mass object. This was quite puzzling unless some collimation was introduced. If GRB emission is not isotropic, the beaming angle may be determined thanks to the presence of achromatic breaks in the afterglow light-curves. These beaming breaks have to be achromatic, but we will see in the next chapter that Swift observations challenge the achromaticity of most of the breaks observed in GRB afterglows. Thus, the determination of the jet angle cannot be considered as a solved problem. Another prediction of jetted GRBs is the existence of orphan afterglows. The frequency of these afterglows could provide another way of measuring the beaming of GRBs. Unfortunately, for the moment these searches have found no convincing candidate.

Another very interesting result was the possible association of GRB 980425 with a supernova, SN 1998bw, which was the subject of much discussion, but is finally confirmed. Several others, discovered with HETE-2 and Swift, were to follow it. The presence of bumps in the late optical afterglows of several GRBs was used to confirm their association with supernovae. The GRB-SN association had important consequences for the nature of the long GRB progenitors, providing strong support to the collapsar model. We will come back in Chapter 8 on this association and on its consequences for the origin of long GRBs.

We have also seen that the first BeppoSAX observations of the afterglows validated the most simple model, the so-called standard model. This model considers fireballs undergoing internal and external shocks, which respectively explain the prompt and the afterglow emissions. With Swift, the standard model encounters difficulties connected with detailed observations of early afterglows, which were not accessible with BeppoSAX because the time needed to point the satellite in the direction of the GRBs found by the WFCs was several hours (5 to 8 h at best).

We have also reported the following results of BeppoSAX:

- The presence of emission and absorption features in the X-ray spectra of some GRBs, and in their afterglows. Such signatures might be quite interesting, but Swift observations have not confirmed the lines detected in X-ray afterglow spectra.
- The discovery of X-ray-rich GRBs. Much more data was collected with HETE-2 on these particularly soft GRBs ($E_p < 50$ keV).
- Preliminary results on the connection between the prompt emission and the afterglow. These results would, however, be largely superseded by Swift, which observes the early X-ray afterglow seconds to minutes after the GRB.
- The last point that we have mentioned is the discovery of a tight correlation between $E_{\text{peak}}$ and $E_{\text{iso}}$, the isotropic equivalent energy radiated by the GRB during its prompt phase. This Amati correlation was the starting point for many discussions as it opened the door to the ‘standardization of GRBs’, that is the possibility of inferring their intrinsic luminosity independent of their redshift.

A final remark, which is perhaps evident, concerns the attribution to BeppoSAX of all the beautiful and fundamental results reported in this chapter. It is fair to mention that follow-up multi-wavelength observations have played a decisive role in the discovery of optical and radio afterglows, in the determination of precise localiza-
tions, in the measurement of redshifts, in the study of GRB host galaxies, and in the characterization of the circumburst medium. Satellite missions are needed to discover GRBs and measure their prompt high-energy emission, but follow-up from the ground is absolutely essential. Without ground-based observations, GRBs would have remained a mystery, and the GRB distance scale would still be unknown. GRB studies provide one of the best illustrations of the impact of multi-wavelength searches in astrophysics. Of course this complementary work was also necessary for GRB missions after BeppoSAX: HETE-2, INTEGRAL, Swift and those still to come.

3.12 REFERENCES


HETE-2 and Swift

4.1 THE HETE-2 MISSION

HETE was the first space mission entirely dedicated to gamma-ray bursts. The concept of a satellite capable of multi-wavelength observations of GRBs was discussed for the first time in 1983, during a GRB meeting held in Santa Cruz (Chupp et al. 1984). It was already evident at that time that GRBs were very elusive events with no counterparts detectable one month after the event. The first realistic implementation of the HETE concept was proposed in 1986, by an MIT-led International Team. This concept emphasized accurate locations and multi-wavelength coverage as the primary scientific goals for a sharply focused small satellite mission. The design and construction of HETE began in 1992, when the program was funded by NASA in the USA, by RIKEN in Japan, and by CNES in France. HETE-1 carried a suite of three instruments:

- four wide-field gamma-ray detectors (6–400 keV), supplied by the CESR of Toulouse, France;
- a wide-field coded-aperture X-ray imager (2–25 keV), supplied by a collaboration including the Institute of Chemistry and Physics (RIKEN) of Tokyo, Japan, and Los Alamos National Laboratory (USA);
- four wide-field near-UV CCD cameras, supplied by the Center for Space Research at the Massachusetts Institute of Technology (USA).

The HETE-1 satellite was launched on November 4, 1996, along with the Argentinian satellite SAC-B, on a Pegasus rocket, from Wallops Island, VA, USA. The Pegasus rocket achieved the right orbit, but its third stage failed to release the two satellites. As a result, SAC-B and HETE-1 were unable to function as designed and both died within a day of launch due to lack of solar power. After the tragic fate of HETE-1 and considering the continuing timeliness of GRB science,
NASA agreed to re-fly the HETE satellite, using spare flight hardware from the first satellite. In July 1997, funding for a second HETE satellite was granted in the three countries contributing to the hardware, and the construction of HETE-2 began immediately. HETE-2 presented three major changes with respect to HETE-1: the HETE-2 satellite bus was built entirely at MIT, not in the industry, the orbit evolved into an equatorial orbit, and the UV cameras were replaced by soft X-ray cameras (Figure 4.1).

The last two changes were direct consequences of the launch of BeppoSAX in April 1996. The experience of the BeppoSAX and RXTE satellites, the former in an equatorial orbit and the latter in a 28-degree orbit, made it clear that electrons and protons in the South Atlantic Anomaly would have a profound effect on the observing efficiency and lifetime of HETE-2 X-ray instruments. In 1998, NASA agreed to fly HETE-2 in an equatorial orbit. The detection of many GRBs by the wide-field cameras of BeppoSAX (see Chapter 3) in early 1997 showed that the prompt X-ray emission from GRBs was often significant. As a result, the HETE team decided to remove the UV cameras from the spacecraft, replacing two of them with CCD-based coded-aperture imagers, sensitive to soft X-rays, the soft X-ray cameras or SXC. The other two cameras were replaced with optical CCD cameras serving as star trackers. HETE-2 data were analyzed on-board by powerful processors, allowing the quick

![Image of HETE-2 satellite](image-url)

**Figure 4.1.** The HETE-2 satellite during final tests. *See also* Color section.
detection of GRBs and, for the first time, the on-board determination of their position on the sky. With BeppoSAX, one had to wait one hour or more for the transmission of WFC data to the ground, and the determination of the GRB position. With HETE-2, the position was computed on-board, usually in seconds, and subsequently transmitted to a network of VHF receivers on the ground (Figure 4.2). This original technology permitted the distribution of accurate GRB positions within tens of seconds of the trigger. Another feature of HETE-2 was its quasi-permanent anti-solar pointing. Anti-solar pointing ensured that the gamma-ray bursts localized by HETE-2 were always on the night side of the Earth, allowing ground telescopes to observe them as soon as they were alerted. The choice of this pointing strategy was dictated by the need for a high synergy between space and ground for gamma-ray burst studies. Overall, it appears that designing a mission entirely dedicated to gamma-ray bursts leads to choices that differ significantly from more classical astronomy missions. We will see in Section 4.7 that this was also the case for Swift.

A description of HETE-2 and its instrumentation can be found in Ricker et al. (2003), and in a series of papers given at the workshop ‘Gamma-Ray Burst and Afterglow Astronomy 2001: A Workshop Celebrating the First Year of the HETE Mission’, published in American Institute of Physics Conference Proceedings No. 662 (2003). The really new features brought by HETE were the fast alerts to the ground and the broad energy coverage of the prompt emission, extending from 1 keV to 400 keV. The rapid alerts unravelled the previously unexplored early optical
afterglows (Section 4.2) and permitted the high-resolution spectroscopy of some very bright optical afterglows. The broad energy coverage was essential to understand the connection between the X-ray flashes, discovered by BeppoSAX, and the classical gamma-ray bursts (Section 4.5). HETE-2 was successfully launched on October 9, 2000.

4.2 FAST ALERTS IN 2002

4.2.1 GRB 021004: optical follow-up from 193s to 3 months after the burst

The first alerts generated on-board HETE-2 were sent in 2002. On October 4, 2002, GCN1565, containing the position of GRB 021004, was distributed to the world 48 seconds after the GRB (Shirasaki et al. 2002), allowing the detection of a bright visible afterglow in the next minutes ($R = 15.3$; Fox 2002). The brightness of the visible afterglow favored its continuous monitoring from 200 seconds to several months after the burst (Fox et al. 2003b, Bersier et al. 2003, Pandey et al. 2003, Holland et al. 2003, Uemura et al. 2003, Fynbo et al. 2005). It also permitted very detailed observations in the visible: broadband photometry (Bersier et al. 2003, Pandey et al. 2003b, Holland et al. 2003), spectroscopy (Möller et al. 2002, Mirabal et al. 2003), spectro-polarimetry (Covino et al. 2002a,b, Rol et al. 2002, Wang et al. 2002, Rol et al. 2003), time-resolved spectroscopy (Matheson et al. 2003, Schaefer et al. 2003, Starling et al. 2003), high-resolution spectroscopy (Savaglio et al. 2002, Fiore et al. 2003), and accurate astrometry (Henden & Levine 2002). Finally, the afterglow was also observed outside the visible domain, in X-rays (Sako & Harrison 2002a,b, Butler et al. 2003b, Fox et al. 2003b), in the infrared (Di Paola et al. 2002, Rhoads, Burud, & Fruchter 2002, Giannini et al. 2002), and at sub-millimeter (Bremer & Castro-Tirado 2002, Kemp et al. 2002, Wouterloot et al. 2002, Smith et al. 2005a) and radio wavelengths (Frail & Berger 2002, Berger, Frail, & Kulkarni 2002, Berger, Kulkarni, & Frail 2002, Pooley 2002a,b,c). The GRB Coordinates Network demonstrated its capabilities on this occasion, with 62 circulars published in the 10 days following GRB 021004.

GRB 021004 was a typical long GRB at redshift $z = 2.326$ (Chornock & Filipenko 2002). Its $\nu F_\nu$ spectrum peaked at 80 keV in our reference frame (270 keV in its rest-frame), and it was moderately bright and soft, with fluences equal to $8 \times 10^{-7}$ and $18 \times 10^{-7}$ erg/cm$^2$ in the energy ranges 2–30 keV, and 30–400 keV, respectively (Sakamoto et al. 2005). These very detailed observations, starting early after the trigger, led to interesting results and conclusions.

The first result was the remarkable evolution of the optical light-curve, which was not smooth, as expected from standard afterglow models, but displayed short-term variability on a timescale of hours, and bumps on a timescale of several days. This variability was accompanied by variations of the optical polarization (Rol et al. 2003, Lazzati et al. 2003) and by color variations (Bersier et al. 2003, Matheson et al. 2003). While Li and Chevalier (2003) showed that the overall evolution of the afterglow at optical, radio, and X-ray wavelengths was fully compatible with the predictions of
wind-interaction models, several authors have proposed different explanations for the origin of this variability. Three explanations have been proposed for the bumps (e.g. Piran 2003): density fluctuations in the medium surrounding the source (Lazzati et al. 2002), episodes of energy injection (Fox et al. 2003b, Björnsson, Gudmundsson, & Jóhanneson 2004, de Ugarte Postigo et al. 2005), or a patchy shell in which the variability is due to random angular fluctuations in the relativistic jet energy (Nakar, Piran, & Granot 2003). Fox et al. (2003b) argued that the energy injection scenario was favored by the fact that X-ray and optical light-curves followed similar trends. The density fluctuation scenario was disfavored because it predicted different behaviors of the optical afterglow and X-ray afterglows: as optical wavelengths are below the cooling frequency, the optical afterglow is sensitive to density fluctuations, while this is not the case for the X-ray afterglow which is above the cooling frequency. Heyl and Perna (2003) re-analyzed optical and X-ray data, and reached the opposite conclusion, favoring density fluctuations. These authors stressed the need for simultaneous observations below and above the cooling frequency to understand the complex behavior of early afterglows (see also Nakar et al. 2003). Finally, Kobayashi and Zhang (2003) attributed the initial bright optical emission to the reverse shock. Dado, Dar, and De Rújula (2004) analyzed the data within the framework of a different model, the cannonball model (see Section 5.3.7), and showed that this model could explain the observations of GRB 021004, if one assumes that two cannonballs were emitted in the explosion.

The second result, indirectly connected with the brightness of the optical afterglow, was the rich spectral information available from the afterglow. High-resolution spectroscopy revealed the existence of several absorption lines blueshifted with respect to the burst, with velocities ranging from 500 to 3000 km s\(^{-1}\). These large velocities, all blueshifted, strongly suggested that the absorbers were physically connected with the burster, and were most probably due to high-velocity material ejected by the star before the explosion. This material could have been steadily ionized by the massive star, and flash ionized by the GRB. There was a general consensus to attribute these absorption features to the fossil stellar wind of a highly evolved Wolf–Rayet star (Møller et al. 2003, Mirabal et al. 2003, Starling et al. 2005). Van Marle, Langer, and García-Segura (2005, 2007) showed, however, that standard models of Wolf–Rayet evolution have difficulties explaining the intermediate velocity components, suggesting that some peculiar evolution is required to produce gamma-ray bursts. Lazzati et al. (2006) have compared the spectra measured at varying times after the GRB, with the prediction of their time-dependent photo-ionization code. They modeled the location and mass of the absorbers, and concluded that the mass loss rate of the progenitor must have been \(\sim 10^{-4} M_\odot \text{yr}^{-1}\), suggestive of a very massive star. Their results confirmed similar conclusions obtained by Schaefer et al. (2003), from a simpler analysis of the spectral lines. Fiore et al. (2005) presented a high-resolution spectrum obtained with UVES on the VLT, and showed how such observations could be used to provide insight into distant galaxies hosting GRBs.

By way of conclusion, we will make two comments. First, we note that in the case of GRB 021004, the inhomogeneity of the medium immediately surrounding the burster seems to be confirmed both photometrically and spectroscopically. Second,
it is quite remarkable that the spectroscopy of the afterglow of GRB 021004 allows the
tomography of parsec-size regions surrounding the burster at a distance of
5.7 Gpc.

4.2.2 GRB 021211: a dark GRB caught early

GRB 021211 was a single-peak, moderately bright burst, lasting 2.5 s in the
30–400 keV energy range, showing clear hard-to-soft evolution, and having a peak
energy of 56 keV (Crew et al. 2002). It was localized by the two imagers of HETE-2,
the WXM and the SXC. The WXM localization was distributed automatically to the
ground 22 s after the trigger, and the much more precise localization of the SXC
(2′ radius vs 14′ for the WXM) was distributed 2 h and 11 min later, after being
checked on the ground (Crew et al. 2002, 2003a). The discovery of the afterglow was
reported by Fox et al. (2002, 2003a), about 53 minutes after the trigger. While GRB
021211 was brighter than GRB 021004 when the first observation was performed
\( R = 14.1 \), 90 s after the trigger (Wozniak et al. 2002, Vestrand et al. 2004), it was 40
times fainter 1 day after the burst \( R > 23.2 \) for GRB 021211—McLeod et al. 2002—
vs \( R = 19.3 \) for GRB 021004—Bersier et al. 2002). The faintness of the optical
afterglow after one day and its brightness in the first minutes showed that some
‘dark bursts’ may just be ‘dim bursts’, stressing the importance of observing very
early after the trigger. The redshift of GRB 021211 measured at the VLT was
\( z = 1.006 \) (Vreeswijk et al. 2006). This moderate redshift made possible the detection
of a supernova culminating about 20 days after the burst (Della Valle et al. 2003).

The optical afterglow of GRB 021211 started with a fast-decay phase lasting
about 10 min (with a decay index \( \alpha_1 = -1.82 \); Li et al. 2003), followed by a slower
decay lasting about 10 days (\( \alpha_2 = -1.1 \)). To the surprise of all the observers, the
optical afterglow of GRB 021211 was incredibly similar to the afterglow of GRB
990123, the only burst for which early afterglow data were available in the pre-HETE
epoch (Fox et al. 2003a, Li et al. 2003, Pandey et al. 2003a, Wei 2003, Holland et al.
2004, Panaitescu & Kumar 2004). This similarity led theoreticians to use for GRB
021211 the recipe which successfully explained the extraordinary GRB 990123. The
early fast-decaying emission was interpreted as being due to the reverse shock and the
late-time afterglow as being due to the forward external shock. This explanation
requires, however, some tuning: Fox et al. (2003a) showed that the radio emission
was fainter than expected in this model, and various authors showed that this inter-
pretation could only explain the observations if the ejecta were strongly magnetized
(Zhang, Kobayashi, & Mészáros 2003, Kumar & Panaitescu 2003, Fan, Wei, &
Zhang 2004, Panaitescu & Kumar 2004, Nysewander et al. 2006). Finally, Dado,
Dar, and De Rújula (2003) explained the optical afterglow of GRB 021211 with the
radiation of a single cannonball.

GRB 021004 and GRB 021211 illustrate the advances permitted by the system-
atic exploration of the early afterglow at optical wavelengths. The possibility of
observing the afterglow so early revealed the complex behavior of early optical
afterglows. This would be confirmed later with more HETE GRBs and with several
Swift GRBs whose optical afterglows exhibit a large variety of behaviors in the
minutes following the burst. In addition, the early characterization of the afterglow, particularly its brightness, provided the opportunity to quickly schedule high-resolution spectroscopic observations of bright afterglows. This was done for the first time for GRB 020813 (Savaglio & Fall 2004, Fiore et al. 2005) and for GRB 021004 (Fiore et al. 2005).

One fascinating aspect of GRB studies is the fact that the next GRB can arrive at any time, for instance while you are looking at the sky or at your computer to monitor the status of your detectors in space. The diversity of GRBs is such that a single one of them, with remarkable properties, can dramatically modify or improve our understanding. That was for instance the case for GRB 970228, GRB 970508, and GRB 980425 for BeppoSAX. It was also the case for GRB 030329 and GRB 050709 for HETE. In the next sections we discuss these two Rosetta-stone GRBs that HETE was lucky enough to localize.

### 4.3 GRB 030329, THE GRB AND THE SUPERNova

On March 29, 2003, HETE detected and localized GRB 030329, the brightest burst of the mission (Vanderspek et al. 2003). It was a double-peak GRB, lasting about 30 s, with a fluence $S_\gamma \sim 1.2 \times 10^{-4}$ erg cm$^{-2}$ (Figure 4.3; Vanderspek et al. 2004). GRB 030329 was so bright that it provoked a ionospheric disturbance (Price et al. 2003, Maeda et al. 2005). About 2 h after the burst, its afterglow was identified as a bright star of 13th magnitude (Peterson & Price 2003) and subsequently observed by the largest telescopes in the world and by a number of instruments of smaller size. At 17 h after the burst, spectroscopy with the ESO’s Very Large Telescope showed that the distance of the source was only 1 gigaparsec ($z = 0.1685$), situating GRB 030329 as one of the closest gamma-ray bursts (Greiner et al. 2003a). At this distance the isotropic equivalent energy released in gamma-rays is $2 \times 10^{52}$ erg. The quick detection and localization of this burst and its proximity had two important consequences: the afterglow of GRB 030329 is among the best-studied afterglows, and GRB 030329 has been unambiguously associated with a known astronomical event: the explosion of a Type Ic supernova. In the case of GRB 030329, the supernova was very energetic, deserving the name of hypernova. We explain below in more detail how these observations deeply affected our understanding of gamma-ray bursts.

**GRB 030329, a well-studied gamma-ray burst**

The quality of the data on GRB 030329 is illustrated in Figure 4.4 (from Lipkin et al. 2004) showing its densely sampled optical afterglow. The brightness of the afterglow made possible its observation with mid-size instruments, and the availability of hundreds of measurements revealed the fine time structure of the afterglow light-curve. The time of the GRB was suitable for observations of the early afterglow from East Asia. The Japanese groups were especially efficient at collecting good photometric measures of the afterglow, with instruments ranging in size from 25 cm to 1 m. The Siding Spring Observatory (Australia) with the 40-inch telescope (Price et al.
Figure 4.3. HETE light-curves of GRB 030329. FREGATE observations are reported from 7 keV to >400 keV, WXM light-curve is measured between 2 and 25 keV and SXC between 2 and 10 keV. The inset figures show more details of small precursors preceding the main peak (Vanderspek et al. 2004).
2003) and ROTSE IIIa (Smith et al. 2003) also provided good measurements of the afterglow in its early phase, and the first GCN Circular relating the discovery of the afterglow (Peterson & Price 2003). After the first hours the afterglow was observed from all around the Earth with optical and radio observatories. We discuss some interesting features of this GRB.

**Optical afterglow**

GRB 030329 triggered HETE at 11:37:14.67 UTC, but the alert was distributed 12:50:24 UT, 73 minutes later. This long delay for a very bright GRB deserves some

**Figure 4.4.** (a) Optical afterglow of GRB 030329 in B, V, R, I bands. For presentation purposes the B V I light curves are shifted vertically by +0.5, −1.5, −1.9 magnitudes respectively (Lipkin et al. 2004). (b) The X-ray to radio energy spectrum of GRB 030329, measured 0.5 days after the burst (Price et al. 2003). Filled circles indicate measurements and open circles extrapolations from measurements at different times. The fit of this spectrum made it possible to determine the properties of the fireball: $v_\infty$, $v_m$, $v_e$, $f_m$, $E$, $n$, $\varepsilon_e$, $\varepsilon_B$, and the electron energy index, $p \sim 2$ (see Chapter 6).
explanation. The $\gamma$-ray detectors of FREGATE, like most gamma-ray detectors, naturally detect charged particles present in space. To avoid triggering FREGATE too often when crossing the high particle fluxes of the South Atlantic Anomaly, the on-board software inhibited triggers when the count rate on the detectors was too high (the zones of high particle background are characterized by very high count rates). Unfortunately it happened that GRB 030329 was so bright that the FREGATE software believed that it was due to particles and inhibited the trigger. After GRB 030329, this feature was disabled in the on-board software and particle trigger inhibition relied only on the Earth longitude of the S/C.

Torii et al. (2003) reported optical observations taken during the burst itself with the Yatsugatake camera in Japan. This ultra-wide-field video system saw no flash brighter than $V = 5.1$ at the time of the GRB. After the distribution of the alert, various narrow-field telescopes observed the afterglow (Sato et al. 2003, Torii et al. 2003, Uemura et al. 2003b, Urata et al. 2004, Burenin et al. 2003). For the first few hours, the optical afterglow had a very smooth light-curve, with no short-term fluctuations, following a power-law decay with slope $\alpha_1 \sim -0.89$, evolving to a steeper decay with $\alpha_2 \sim -1.19$ after 0.25 day. Torii et al. (2003) interpreted the smooth power law as the indication of a smooth medium around the source or as the indication that the optical range was above the cooling frequency (in which case the optical light-curve is insensitive to density fluctuations). Gorosabel et al. (2006) also noted that this lack of fluctuations put strong constraints on the existence of small-scale angular inhomogeneities within the jet. After half a day a second break was clearly discernible, leading to a steeper decay, with $\alpha_3 \sim -1.81$ (Burenin et al. 2003, Sato et al. 2003, Uemura et al. 2003b). Berger et al. (2003) found a clear break in the radio afterglow 10 days after the burst, which they interpret as a jet break from a jet with an opening angle of 17°. An optical break was difficult to detect at that time because the supernova light dominated the emission (see below). Berger et al. (2003) also interpreted the optical break at 0.5 day as a jet break, for a jet having an opening angle of 5°. They concluded that GRB 030329 was the result of a two-component explosion: a narrow (5°) ultra-relativistic component responsible for the gamma-rays and the early afterglow, and a wide, mildly relativistic component responsible for the radio and optical afterglows beyond 1.5 days. In this scenario, the total energy release is dominated by the wide component which is 10 times more energetic than the narrow highly relativistic jet. A re-analysis of the data sets from five observatories including 2687 photometric measurements obtained between 78 min and 79 days after the burst was performed by Lipkin et al. (2004). They show that the afterglow cannot be simply described by a succession of power-law segments but is best fit by a series of broad bumps superimposed on a broken power law with indices $\alpha_1 \sim -1.1$ and $\alpha_2 \sim -2.0$, and a break around 3 to 8 days (see also Uemura et al. 2003b). They found that the break at 0.25 days reported by Torii et al. (2003) was accompanied by a variation of the color of the afterglow, and can be interpreted as a cooling break. According to their analysis, the break at 0.5 days is not a jet break but is part of the general undulations superimposed on the decaying broken power law. Granot, Nakar, and Piran (2003) explained the fluctuations observed in the early afterglow light-curve with refreshed shocks, slow shells ejected by the source catching up with
the ejecta after the jet break. The energy of the refreshed shocks multiplies by 10 the total energy content of GRB 030329, bringing it close to the majority of observed GRBs. Huang, Cheng, and Gao (2006) re-examined this issue with numerical simulations based on an analytical model and showed that the two-component jet model cannot produce the fast rising re-brightening observed at 1.6 day. They concluded that a sequence of energy-injection events provides the best explanation for the re-brightening at \( \sim 1.6 \) days, and for the overall optical light-curve up to \( \sim 80 \) days.

At late times (>1 day), more detailed observations were performed including multi-band photometry, polarimetry, and spectroscopy. Greiner et al. (2003b), in particular, reported the most complete and dense sampling of the polarization behavior of a GRB afterglow, showing a low, varying level of polarization. Such measurements are crucial to confirm the interpretation of breaks in afterglow light-curves and to study the physical properties of dust and magnetic fields in high-z galaxies (Klose et al. 2004). Multi-band photometry showed that the re-brightening episodes at days 1.6 and 5.2 were not associated with color changes, and suggested a low extinction in the host galaxy (Simon, Hudeck, & Pizzichini 2004, Bloom et al. 2004).

**The radio and millimeter afterglow**

The afterglow of GRB 030329 was also extensively observed at radio and millimeter wavelengths (Sheth et al. 2003, Finkelstein et al. 2004, Kuno et al. 2004, Taylor et al. 2004, 2005, Frail et al. 2005, Kohn et al. 2005, Resmi et al. 2005, Smith et al. 2005b, van der Horst et al. 2005, Pihlström et al. 2007). These observations showed a clear break around 10 days, followed by a rapid decay with an index \(-2.0\). As predicted by the standard fireball model, the break arrives earlier at higher frequencies. Sheth et al. (2003) showed that the millimeter afterglow is compatible with a double-jet model proposed by Berger et al. (2003), while Resmi et al. (2005) showed that this model was also supported by the combination of millimeter, visible, and X-ray observations. An achromatic flattening of the radio afterglow at \( T_\text{b} + 50 \) days was interpreted by Frail et al. (2005) as the indication of a transition to sub-relativistic expansion, making it possible to assess the total energy released in the explosion: a few \( 10^{51} \) ergs, similar to values found in other GRBs.

The relative proximity of GRB 030329 allowed measuring its expansion velocity with Very Long Baseline Interferometry (VLBI) observations. Taylor et al. (2004, 2005) found a size of the afterglow which was \( \sim 0.07 \) mas (0.2 pc) 25 days after the burst and 0.17 mas (0.5 pc) 83 days after the burst, indicating an apparent velocity of 3\( c \)–5\( c \).\(^1\) According to Berger et al. (2003), the size of the fireball was 0.02 mas (0.07 pc) 15 days after the burst. This high expansion velocity was consistent with expectations of the standard fireball model. Pihlström et al. (2007) presented high-sensitivity

---

\(^1\)This apparent super-luminal velocity is the consequence of the relativistic motion of the emitting plasma in our direction. As explained by Rees (1966) the apparent transverse velocity is \( \beta_{\text{app}} = \beta \sin(\theta)/(1 - \beta \cos(\theta)) \), where \( \beta \) is the true velocity of the radiating matter, \( \theta \) the angle that our line of sight makes with the motion of the matter, and \( \beta_{\text{app}} \) the apparent velocity. For instance, matter moving towards us with an angle of \( \theta = 10^\circ \) and a (true) velocity \( \beta = 0.95c \) (\( c \) is the speed of light), would appear to have an apparent transverse velocity \( \beta_{\text{app}} = 2.56c \).
VLBI observations 806 days after the GRB giving an angular diameter of the radio afterglow of $0.347 \pm 0.09$ mas ($0.99 \pm 0.26$ pc), showing a slow-down of the apparent expansion velocity.

The X-ray afterglow was observed by RXTE and XMM (Tiengo et al. 2003, 2004), but it was not so densely sampled as the optical and radio afterglows. Overall, the X-ray observations were compatible with the expectations of the fireball model (Willingale et al. 2004). GRB 030329 took place two billion years ago in a blue subluminous galaxy ($M \sim 10^8 M_\odot$, $R = 22.6$), actively forming stars (Gorosabel et al. 2005).

**GRB 030329 and SN 2003dh**

The brightness of the afterglow of GRB 030329 encouraged observers to perform detailed spectroscopic observations in the days following the explosion. As shown in Figure 4.5(a) (from Stanek et al. 2003), in the days following the burst the optical spectrum exhibited a remarkable transition from a featureless power-law spectrum characteristic of GRB afterglows to the typical spectrum of a hypernova, which was called SN 2003dh (Hjorth et al. 2003, Matheson et al. 2003, Stanek et al. 2003, Kurt et al. 2005). These observations permitted characterizing the supernova—clearly a hypernova as explained below—and measuring the time of its explosion. They securely associated long GRBs with the deaths of massive stars, as predicted by the collapsar model (see Chapter 8).

The first spectroscopic observations were performed with the BTA-SAO RAS 6-m telescope 11 h after the burst. Curiously the spectra taken at this very early time already displayed broad lines typical of supernovae that would become predominant in spectra taken several days after the GRB (Kurt et al. 2005). The spectrum of the OT + supernova was observed regularly during the month of April with the ESO-VLT (Hjorth et al. 2003), with the Multiple Mirror Telescope, the 1.5-m Tillinghast telescope at the Fred Lawrence Whipple Observatory, and the Magellan 6.5-m Clay telescope (Stanek et al. 2003), with the du Pont 2.5-m telescope at LCO, the Shane 3-m telescope at Lick Observatory, and the Keck I and II 10-m telescopes (Matheson et al. 2003), and with Subaru in May and June (Kawabata et al. 2003, Kosugi et al. 2004). The combination of these spectra with photometric data permitted separating the afterglow and SN contributions. The light of the supernova started to be visible 7 days after the burst and dominated the emission after 11 days (Matheson et al. 2003). The light-curve of the supernova and its spectrum resembled those of 1998bw, the hypernova associated with GRB 980425, and 1997ef, a less energetic hypernova. The timing of the supernova indicated that it exploded within two days of the GRB. Around maximum the spectra of SN 2003dh resemble those of SN 1998bw (see Figure 4.5(b)), but at later times they looked more like those of the less energetic hypernova SN 1997ef.

Mazzali et al. (2003) modeled the spectra of SN 2003dh using radiation transport codes. They showed that SN 2003dh had a high explosion kinetic energy ($\sim 4 \times 10^{52}$ erg in spherical symmetry), making it one of the most powerful hypernovae observed so far. However, the light-curve derived from fitting the spectra
Figure 4.5. (a) Evolution of the GRB 030329–SN 2003dh optical spectrum from 2.64 days after the burst (April 1.13 UT) to 9.64 days after the burst (April 8.13 UT). Early spectra consist of a power-law continuum ($I_{\nu} \propto \nu^{-0.9}$) with narrow emission lines originating from HII regions in the host galaxy at redshift of $z \sim 0.168$. Spectra taken at later times, after April 5, show broad peaks characteristic of a supernova (Stanek et al. 2003). (b) MMT spectrum measured on April 8 with continuum subtracted. This spectrum is similar to the spectrum of the peculiar type Ic SN 1998bw a week before maximum light. The match is not as good for SN 1998bw at maximum light, especially at the red end of the spectrum (Stanek et al. 2003).
suggests that SN 2003dh was not as bright as SN 1998bw, ejecting only \( \sim 0.35 M_\odot \) of \(^{56}\text{Ni}\). The mass of the ejecta is \( \sim 8 M_\odot \), somewhat less than in the other two hypernovae (1998bw and 1997ef). The progenitor must have been a massive star \((M \sim 35–40 M_\odot)\), as for other hypernovae. The need to combine different one-dimensional explosion models strongly indicates that SN 2003dh was an asymmetric explosion. Deng et al. (2005) found similar numbers. The spectra can be reproduced with a spherical ejecta model of \( M_{ej} \sim 7 \pm 3 M_\odot \), \( E_K \sim 3.5 \pm 1.5 \times 10^{52} \text{erg} \), with \( E_K/M_{ej} \sim 5 \), and \( M(^{56}\text{Ni}) \sim 0.4^{+0.15}_{-0.1} M_\odot \). These numbers suggest a progenitor main-sequence mass of \( \sim 25–40 M_\odot \), significantly higher than normal Type Ic SNe. Kawabata et al. (2003) and Kosugi et al. (2004) have studied the spectrum of the supernova measured with the Subaru 8.2-m telescope at later times, 40 and 85 days after the burst. The second observation took place during the transition between the photospheric and the nebular phases. Their analysis of the nebular-phase emission lines suggests that the explosion of the progenitor of GRB 030329 was aspherical, and that the axis of asphericity was aligned with the GRB relativistic jet and with our line of sight. Although the decay index and the color evolution in the rest-frame during the two epochs were consistent with those of the Type Ic supernovae SN 1998bw and SN 1997ef, they found that the nebular-phase lines emerged slightly earlier in 2003dh than in the case of these supernovae.

A new understanding of gamma-ray bursts was reached with the discovery of their close connection with supernova explosions, a well known astronomical phenomenon. It became clear that massive collapsing stars could lead to two of the most remarkable astronomical events: hypernovae and gamma-ray bursts. At the same time exciting questions emerged concerning the origin of the difference between normal Type Ic supernovae and GRB-emitting supernovae and the nature of the central engine producing the burst, which was suspected to be a fast-spinning black hole but which could also be a newly born magnetar. These issues are discussed from a theoretical point of view in Chapter 8.

### 4.4 GRB 050709, THE FIRST OPTICAL AFTERGLOW OF A SHORT GRB

On July 9, 2005, at 22:36:37 UT, the Soft X-Ray Camera (SXC), the Wide-Field X-Ray Monitor (WXM) and the French Gamma Telescope (FREGATE) instruments on-board the HETE-2 satellite detected GRB 050709, a short-hard GRB lasting 0.07 s, followed by a long-soft bump from the same location (Butler et al. 2005a, Villasenor et al. 2005; Figure 4.6). HETE localized the burst in a circular region with a radius of 81 arcsec. This region was observed with NASA’s Chandra X-ray Observatory 2.5 and 16 days after the burst. In the first epoch image a bright point-like source was detected, which had faded by more than a factor of 2 in the second epoch image (Fox et al. 2005). In the meantime optical observations at the Danish 1.5-m telescope at ESO’s La Silla Observatory showed an optical transient which disappeared after one week (Hjorth et al. 2005). This is the typical behavior of GRB afterglows. Interestingly, the burst was localized in the outskirts of an irregular
Figure 4.6. (A) Gamma-ray light curve of GRB 050709 showing the short GRB followed by a long, soft bump. The association of the two allowed identification of the first optical afterglow of a short GRB. WXMM observations are reported in panels (a) and (b). FREGATE observations are shown in panels (c) and (d). The short, hard spike has a duration $T_{90} = 200 \pm 50$ ms in the 2–25 keV and 70 ± 10 ms in the 30–400 keV energy band. The long, soft bump has a duration $T_{90} = 130 \pm 7$ s in the 2–25 keV energy band (Villasenor et al. 2005).

Figure 4.6. (B) These Hubble Space Telescope images show the fading afterglow and host galaxy of GRB 050709. The images are taken 5.6, 9.8, 18.6, and 34.7 days after the burst, respectively. The bright, point-like afterglow is located to the left, and fades away over the course of the month following the burst. The colors indicate the intensity of red light (814 nm) as seen by the Advanced Camera for Surveys instrument on HST (Fox et al. 2005). See also Color section.
star-forming galaxy at redshift $z = 0.16$, which is certainly its host (see Figure 4.6). A few days after this discovery, Swift localized another short GRB, GRB 050724, that also showed an optical afterglow (Berger et al. 2005). In that summer of 2005, it was as if Nature had decided that the mystery of the origin of short GRBs had lasted long enough. But the truth was that the technology was finally ripe to allow us to explore a new class of high-energy cosmic transients.

While the existence of two classes of GRBs had been known for a long time, the considerable advances made possible by the discovery of the afterglows concerned only long GRBs with durations longer than a few seconds. At the beginning of 2005, the afterglows of short GRBs remained unobserved, and their origin mysterious. The main reason for this situation was that short GRBs, lasting tens to hundreds of milliseconds, contained far fewer photons than long GRBs and were consequently much harder to localize with coded mask imagers. Fortunately, Nature helped us with the emission of long-lasting ($\sim 100$ s) soft bumps following some short GRBs. The number of photons contained in these soft bumps can be considerably larger than the number of photons in the short burst itself, providing a useful signal for the localization of these events with low-energy detectors. Hints of this long-lasting emission had been previously observed in short GRBs detected with BATSE (Lazzati, Ramirez-Ruiz, & Ghisellini 2001), Konus (Frederiks et al. 2004) and BeppoSAX (Montanari et al. 2005), but its potential for the localization of short GRBs was not realized. The identification of the optical afterglows of GRB 050709 and GRB 050724 revealed their similarity with long GRBs: both types exhibit a burst of high-energy photons followed by radio, optical, and X-ray afterglows, a strong indication that both types of events are due to ultra-relativistic jets pointing in our direction. Yet, the discovery of the afterglows of short GRBs revealed many significant differences with their longer cousins. While GRB 050709 and GRB 050724 lie close to their host galaxies, in contrast to long GRBs they did not occur in star-forming regions (Barthelmy et al. 2005b). In fact the host of GRB 050724 is an elliptical galaxy with no on-going star formation (Berger et al. 2005b). Also, at redshift $z = 0.1–0.2$, they are on average 10 times closer than long GRBs and release $10^2$ to $10^3$ times less energy. A third very interesting difference is that short GRBs are not associated with supernovae (Covino et al. 2006, Fox et al. 2005, Hjorth et al. 2005, Watson et al. 2006) while nearby long GRBs are most often associated with bright supernovae (see, however, Section 4.10.4). All these differences show that short and long GRBs have distinct progenitors. The properties of the rare short bursts with an optical afterglow strengthen models which associate them with mergers of binary compact objects containing a black hole plus a neutron star, or two neutron stars (see Chapter 8). After the discovery of the optical afterglows of GRB 050709 and GRB 050724, more short GRBs have been detected and localized with Swift; they are discussed in Section 4.11 at the end of this chapter.

**A target for gravitational-wave detectors**

The association of short GRBs with coalescing compact stars in binary systems, and their relative proximity (with respect to long GRBs) has interesting consequences for
the detection of gravitational waves. We now have the confirmation that in some cases the coalescence of two stars can be accompanied by a strong electromagnetic signal, a short GRB, which tells us the time and direction of the event. This signal may be particularly helpful to identify the signature of astrophysical events in the huge amount of data produced by current gravitational wave detectors like VIRGO or LIGO, and future detectors like LISA. This is only one of the fascinating perspectives opened by the detection of the first optical afterglow from a short GRB (Abbott et al. 2008).

4.5 X-Ray Flashes, The GRB Soft Cousins

4.5.1 The discovery of X-ray flashes

The discovery of GRBs was enabled by the use in space of detectors sensitive to non-thermal $\gamma$-rays (see Chapter 1). While sources of thermal X-rays (up to several kiloelectronvolts) are very numerous in the sky (especially in the Galaxy), sources of non-thermal $\gamma$-rays are much less frequent, making the detection of GRBs quite easy above 30 keV. Below 30 keV the detection of GRBs is made much more difficult by the increased level of diffuse background and by the existence of an enormous background of bursting X-ray sources in the Galaxy. Given the large fields of view of GRB detectors, it is almost impossible for non-localizing detectors detecting a soft burst to be sure that it is not due to a galactic source. One could rely on the energy spectrum of the transient to infer its nature, but soft events are only detected between a few keV and 10–20 keV and it is difficult to measure the difference between a thermal and a non-thermal spectrum in such a small energy range, especially when the direction of the burst is not known.

Evidence of the existence of soft GRBs was brought by the Japanese Ginga spacecraft which detected some very soft bursts (Strohmayer et al. 1998). Ginga was equipped with a proportional counter allowing it to detect X-ray photons down to 2 keV. Despite the fact that Ginga and BATSE triggered in the same energy range (50–300 keV), Ginga found some GRBs for which the ratio of the energy emitted in X-rays (2–10 keV) relative to $\gamma$-rays (50–300 keV) was much larger than the usual value of a few percent and sometimes larger than unity. The existence of soft GRBs was neatly confirmed by BeppoSAX with the discovery of X-ray flashes. We have seen in Chapter 3 that 10% of the GRBs detected by the GRBM were in the field of view of the Wide Field Camera. After some time, Heise et al. (2001) who had seen many GRBs in the WFC, realized that some events seen in the WFC, which were not detected by the GRBM and not localized to known X-ray bursters, looked exactly like GRBs. They proposed an operational definition of X-ray flashes (XRFs) seen in the WFC, namely: ‘A Fast Transient X-ray source with duration less than 1000 sec which is not triggered and not detected by the Gamma-Ray Burst Monitor in the $\gamma$-ray range 40–700 keV’. In nearly 5 years of operation (1996–2001) the WFC had detected 17 XRFs corresponding to this definition, and 49 GRB counterparts. Since BATSE was still in operation at that time, Heise et al. asked the BATSE team to look
in their data if they could find the WFC events. Five events were not detectable by BATSE because they were occulted by the Earth. None of the twelve remaining events triggered BATSE, but a careful look at the continuous data available for ten events revealed that nine of them were seen in the low-energy range of BATSE (20–100 keV) with significance larger than 5σ (Kippen et al. 2001). The comparison of the properties of these nine events with the full sample of BATSE GRBs showed that they had similar durations, that they followed a hard-to-soft evolution and that they appeared generally consistent with the extrapolated hardness–intensity trend of long-duration bursts. Kippen et al. (2001) concluded that XRFs represent a previously unexplored part of the GRB population. They confirmed this finding with a study of the broad-band spectra of 10 XRFs and 18 GRBs seen by both the BATSE and the WFC (Kippen et al. 2003).

The question of the origin of XRFs was first tackled by Heise et al. (2001, see also Heise 2003) who suggested several possibilities, all linked with GRBs: (1) highly redshifted GRBs, (2) GRBs seen from the side of the jet, and (3) GRBs with low Lorentz factors, possibly associated with dirty fireballs. Zhang and Mészáros (2002) discussed these explanations from a theoretical point of view. They showed that the dirty fireball model is not favored because dirty fireballs have closer-in internal shock radii, higher magnetic fields and thus higher $E_{\text{peak}}$. They also noted that XRFs, while being soft, do not show additional indications of high redshift ($z > 6$), like time dilation. Yamazaki, Ioka, and Nakamura (2002, 2003) studied the off-axis model in some detail and found that XRFs could be explained by off-axis GRBs at low redshifts ($z < 0.2$). Bloom et al. (2003) presented Chandra, Hubble Space Telescope (HST), and Keck observations of the fields of two XRFs with X-ray afterglows, XRF 011030 and XRF 020427 (and a radio afterglow for XRF 011030). They showed that these two XRFs were associated with faint blue galaxies at redshift smaller than $z \sim 3.5$, thus excluding the high redshift hypothesis.

4.5.2 HETE-2 confirms that X-ray flashes are soft GRBs

With its broad energy range, HETE-2 was able to trigger on X-ray flashes as well as on GRBs. The significant overlap of the energy ranges of FREGATE (6–400 keV) and WXRT (2–20 keV) made it possible to detect GRBs with a broad range of peak energies. The spectral properties of HETE-2 XRFs have been discussed by Barraud et al. (2003) and Sakamoto et al. (2005). The prompt emission of some remarkable XRFs is discussed in Amati et al. 2004 (XRF 020427—BeppoSAX), Sakamoto et al. 2004 (XRF 020903—HETE), Romano et al. 2006, and Sakamoto et al. 2006 (XRF 050416—Swift), Arimoto et al. 2007 (XRF 040916—HETE), Stratta et al. 2007 (XRF 040912—HETE), Sakamoto et al. 2006 (XRF 050416A—Swift), Campana et al. 2006a (XRF 060218—Swift). Barraud et al. (2003) and Sakamoto et al. (2005) studied the properties of the XRF prompt emission and showed that they form an extension of the GRB population towards lower energy events. Sakamoto et al. (2005) found that the HETE sample contains about 1/3 XRFs, 1/3 GRBs, and 1/3 intermediate events called X-ray-rich GRBs. They proposed an operational definition of these events based on the fluence ratio above and below 30 keV: bursts with $S_{30–400}/S_{2–30} < 0.3$ are called GRBs, bursts with $0.3 < S_{30–400}/S_{2–30} < 1$
are called X-ray-rich GRBs, and bursts with \( S_{\text{[2-20]}} / S_{\text{[30-400]}} > 1 \) are called XRFs. They finally showed that XRFs and GRBs differ only by their \( E_{\text{peak}} \), while they have similar durations and spectral indices.

The fast localization of X-ray flashes with BeppoSAX (at the end of the mission), HETE-2 and Swift, provided considerably more insight on these events. We were eventually able to study their afterglows at various wavelengths, their supernovae, and their host galaxies, and to measure their distances. XRF 020903 provides an excellent example of the information accumulated on XRFs in the year following the discovery of these events. With \( E_{\text{p}} \leq 4 \text{ keV} \), a duration of about 10 s, and hard-to-soft evolution during the burst, XRF 020903 is undoubtedly an X-ray flash (Sakamoto et al. 2004). Its faint optical afterglow \( (R = 19.3 \text{ at maximum, } 0.9 \text{ day after the burst}) \) was superimposed on a relatively bright galaxy \( (R = 20.9) \), and it could only be detected after several days, using image subtraction methods (Soderberg, Kulkarni, & Frail 2003). The burst took place in a galaxy, which is part of a complex association at redshift \( z = 0.25 \) (Soderberg et al. 2004a, 2005, Bersier et al. 2006). Given this distance, the energy released at high-energy during the prompt phase was \( 2 \times 10^{49} \text{ erg} \), intermediate between the ‘anomalous’ GRB 980425 \( (10^{48} \text{ erg, } z = 0.0085) \) and more classical GRBs detected at redshift \( z \sim 1 \). Late-time observations with HST permitted subtracting accurately the magnitude of the host galaxy and measuring the evolution of the brightness of the afterglow. The decay of the afterglow shows a clear plateau around \( 10-30 \text{ days} \), which has been interpreted as the emergence of a supernova like 1998bw, but 0.8 magnitude dimmer (Soderberg et al. 2005, Bersier et al. 2006). Another well studied X-ray flash is XRF050416A (Sakamoto et al. 2006, Holland et al. 2007, Mangano et al. 2007, Soderberg et al. 2007). Detailed observations of this event have shown that it shares the main characteristics of GRBs. XRF050416A took place in a star-forming galaxy at redshift \( z = 0.65 \), a bump in the light-curve of its optical afterglow can be attributed to a supernova with peak optical luminosity roughly comparable to that of SN 1998bw, and the light-curve of its X-ray afterglow follows the canonical shape observed for GRBs (see Section 4.9.1). XRF 050416A exhibits nevertheless an unusual feature, a bright radio flare 40 days after the burst (Soderberg et al. 2007).

Since XRFs are intrinsically fainter than GRBs, they are preferentially detected at smaller distances (see below) and they may be privileged candidates for the detection of associated supernovae. As we have seen, evidence for supernova light has been found in the optical afterglows of XRF 020903 and XRF 050416A. Supernovae have been searched for with HST in various other XRFs (Levan et al. 2005, Soderberg et al. 2005), with no positive detection. Very strong limits were obtained for XRF 040701 at \( z = 0.21 \), but it is difficult to draw firm conclusions from this non-detection because XRF 040701 was a dark GRB which could be heavily extinguished.

The redshifts of several XRFs have been measured: XRF 020903 at \( z = 0.25 \) (Soderberg et al. 2004a), XRF 030429 at \( z = 2.66 \) (Jakobsson et al. 2004), XRF 030528 at \( z = 0.782 \) (Rau, Salvato, & Greiner 2005), XRF 040701 at \( z = 0.21 \) (Kelson et al. 2004), XRF 040912 at \( z = 1.56 \) (Stratta et al. 2007), XRF 050408 at \( z = 1.236 \) (Berger, Giadlers, and Oemler 2005), XRF 050416A at redshift \( z = 0.65 \) (Cenko et al. 2005), XRF 050824 at \( z = 0.83 \) (Sollerman et al. 2007), XRF 060218 at \( z = 0.033 \).
(Mirabal & Halpern 2006). These redshifts demonstrate that XRFs are not highly redshifted GRBs. Quite remarkably, all these events appear compatible with the $E_{\text{peak}} - E_{\text{iso}}$ relation derived by Amati et al. (2002) for gamma-ray bursts. The spectral and temporal properties of XRFs, their range of redshifts, the nature of their host galaxies, and the detection of supernovae associated with some of them suggest a common physical origin with GRBs.

### 4.5.3 The nature of X-ray flashes

After 2004, various models were proposed to explain X-ray flashes, all based on variations of the ‘standard’ GRB model. The low $E_{\text{peak}}$ of XRFs has been explained by a different look at standard GRBs (e.g. observation at large angles) or by an intrinsic softness. We review these models below.

**XRFs as GRBs viewed off-axis**

Dado, Dar, and De Rújula (2004) analyzed XRFs in the framework of the cannonball model. They concluded that XRFs and GRBs are intrinsically identical objects, but that the XRFs are viewed from angles (relative to the jet direction) that are typically a few times larger than the typical viewing angles of ‘classical’, long-duration GRBs. They predicted the existence of XRFs with durations similar to those of short GRBs (<2 s).

Yamazaki, Ioka, and Nakamura (2004) proposed GRB jets consisting of multiple subjets or subshells (inhomogeneous jet model). If the multiplicity of the subjets along a line of sight, $n_s$, is large ($n_s \gg 1$) the event looks like a long GRB, while if $n_s$ is small ($\sim 1$) the event looks like a short GRB. If the line of sight is off-axis for all subjets, the event looks like an X-ray flash or an X-ray-rich GRB. The same model explains the lognormal distribution of durations of short and long GRBs.

Granot, Ramirez-Ruiz, and Perna (2005) examined the predictions of the various XRF models for the afterglow emission and tested them against the observations of two events with well monitored afterglow light curves: XRF 030723 and GRB 041006 (an X-ray-rich GRB). They showed that most existing XRF models are hard to reconcile with the observed afterglow light curves, which are very flat at early times. Such light curves are, however, naturally produced by a roughly uniform jet with relatively sharp edges that is viewed off-axis (from outside of the jet aperture). This type of model self-consistently accommodates both the observed prompt emission and the afterglow light curves of GRB 041006 and XRF 030723, implying viewing angles $\theta_{\text{obs}}$ from the jet axis of $(\theta_{\text{obs}} - \theta_0) \sim 0.15\theta_0$ and $(\theta_{\text{obs}} - \theta_0) \sim \theta_0$, respectively, where $\theta_0 \sim 3^\circ$ is the half-opening angle of the jet. This suggests that GRBs, X-ray-rich GRBs, and XRFs are intrinsically similar relativistic jets viewed from different angles. A GRB is observed when $\Gamma(\theta_{\text{obs}} - \theta_0) < 1$, an X-ray-rich GRB is observed when $1 < \Gamma(\theta_{\text{obs}} - \theta_0) < \text{few}$, and an XRF is observed when $\Gamma(\theta_{\text{obs}} - \theta_0) > \text{few}$. $\Gamma$ is the Lorentz factor of the outflow near the edge of the jet, from which most of the observed prompt emission arises.

Jin and Wei (2004) studied numerically the fluxes, spectra and peak energies predicted by off-axis models and found that non-uniform jets are more appropriate
than uniform jets to reproduce these observational properties. They predicted observed ratios of gamma-ray bursts to X-ray flashes of the order of a few. Zhang et al. (2004) showed that the fact that XRFs follow the $E_{\text{peak}} - E_{\text{iso}}$ relation poses problems both for a power-law universal jet model (in which the energy per solid angle decreases as the inverse square of the angle with respect to the jet axis), and for a conical jet model (with a uniform energy density within the jet beam and a sharp energy cut-off at the jet edge). They showed that the GRB-XRF prompt emission/afterglow data can be understood if GRB-XRF jets are quasi-universal and structured, with a Gaussian-like or similar structure. In this picture the jet has a characteristic angle within which there exist mild variations of energy and outside which there is a very rapid (e.g. exponential) decrease of energy. Monte Carlo simulations show that quasi-universal Gaussian jets with typical opening angles of $5.7^{+3.4}_{-2.1}$ degrees and a standard jet energy of about $\log(E_\gamma, \text{erg}) = 51.1 \pm 0.3$ can explain the data. According to this model, the true-to-observed number ratio of the whole GRB + XRF population is about 14.

Lamb, Donaghy, and Graziani (2005) noted that the density of bursts found by HETE-2 is roughly constant per logarithmic interval in burst fluence $S_\gamma$ and observed spectral peak energy $E_{\text{peak}}$. Using population synthesis simulations of the bursts and detailed modeling of the instruments that detect them, they explore a unified jet model producing GRBs, X-ray-rich GRBs and XRFs. They show that both a variable jet opening angle model (in which the emissivity is a constant independent of the angle relative to the jet axis) and a universal jet model (in which the emissivity is a power-law function of the angle relative to the jet axis) can explain the observed properties of GRBs reasonably well. However, if one tries to account for the properties of XRFs, X-ray-rich GRBs, and GRBs in a unified picture, the extra degree of freedom available in the variable jet opening angle model enables it to explain the observations reasonably well whereas the power-law universal jet model cannot. The reason is that the probability of observing intrinsically faint bursts like XRFs is very low in the power-law universal jet model. The variable jet opening angle model of XRFs, X-ray-rich GRBs, and GRBs implies that most GRBs have very small jet opening angles ($\sim 0.5^\circ$) and that the energy $E_\gamma$ radiated in gamma-rays is $\sim 100$ times less than the values found in standard beaming-corrected calculations (see Section 3.4). It also implies that there are $\sim 10^4 - 10^5$ GRBs with very small jet opening angles for every burst that is observable. If this is the case, the rate of GRBs could be comparable to the rate of Type Ic core-collapse supernovae.

D’Alessio, Piro, and Rossi (2006) studied a large sample of XRFs and X-ray-rich GRBs in order to understand their connection with classical GRBs. They found that the spectral properties of the prompt emission are similar to those of classical GRBs, albeit with a peak energy which is lower by a factor of 4. They also found that the X-ray and optical flux distributions and the light-curves of the XRRs/XRFs afterglows are consistent with those of classical GRBs. They concluded that the high-redshift scenario can explain some but not all the events (as discussed in Section 4.4) and that the off-axis model may be consistent with the observations when a homogeneous jet is considered.
Hence, the question of the shape and geometry of GRB jets is far from being resolved (even if energetic considerations tell us that GRBs must be collimated), and we will see now that it is not even sure that X-ray flashes are due to geometric effects. Let us discuss below the models which can produce intrinsically soft GRBs.

**XRFs as intrinsically soft GRBs**

Gendre, Galli, and Piro (2007) have made a systematic analysis of the X-ray afterglows of XRFs with known redshift observed by Swift. They derived their redshift and luminosity distributions, and compared their properties with a sample of normal GRBs observed with the same instrument. They found that the luminosity distribution of XRF afterglows is similar to the one of GRBs, ruling out most off-axis models, but the homogeneous jet model. However, this model is also ruled out as it predicts a GRB rate uncomfortably close to the observed rate of supernovae. They conclude that XRFs in their sample are intrinsically soft events.

In the context of the collapsar model of GRBs, Mizuta et al. (2006) investigated the dynamics of an outflow propagating within the progenitor through two-dimensional axisymmetric relativistic hydrodynamic simulations. They locally inject an outflow near the center of a progenitor and calculate 25 models in total, with fixed total input energy \( (10^{51} \text{ erg s}^{-1}) \) and radius of the injected outflow \( (7 \times 10^7 \text{ cm}) \), and varying bulk Lorentz factor from the relativistic to the non-relativistic regime \( \Gamma_0 = 5–1.05 \), corresponding to \( c/\xi \approx 0.98–0.3 \), with \( c \) being speed of light and a specific internal energy, \( \varepsilon_0/\xi^2 \) vary in the range 0.1–30 \( \Gamma_0 \) and \( \varepsilon_0 \) respectively characterize the kinetic and thermal energy per particle in the outflow. They observe a smooth but dramatic transition from a collimated jet to an expanding outflow among calculated models. The half opening angle of the outflow (\( \theta \)) is sensitive to \( \Gamma_0 \), with for instance \( \theta < 2^\circ \) for \( \Gamma_0 > 3 \). The maximum Lorentz factor is sensitive to both \( \Gamma_0 \) and \( \varepsilon_0 \), roughly \( \Gamma_{\text{max}} \sim \Gamma_0 (1 + \varepsilon_0/\xi^2) \). The jet structure totally depends on \( \Gamma_0 \). When \( \Gamma_0 \) is high, high-pressure progenitor gas heated by the bow shock collimates the outflow to form a narrow, relativistic jet. When \( \Gamma_0 \) is low, on the contrary, the outflow expands soon after the injection, since the bow shock is weak and thus the pressure of the progenitor gas is not high enough to confine the flow. They explain the smooth transition between the GRBs, X-ray-rich GRBs (XRRs), and X-ray flashes (XRFs) by the same model but with different \( \varepsilon_0 \) values.

Barraud et al. (2005) used a simplified model for internal shocks to check if XRFs can be intrinsically soft events due to some specific values of the parameters describing the relativistic outflow emerging from the central engine. They generated a large number of synthetic events and found that XRFs are obtained when the contrast \( \Gamma_{\text{max}}/\Gamma_{\text{min}} \) of the Lorentz factor distribution is small, while the average Lorentz factor \( \Gamma \) is large. They excluded the possibility that the bulk of the XRF population were GRBs at large redshift and showed that a large population of XRFs could be explained if outflows with a small contrast are commonly produced. If conversely the Lorentz factor distribution within the wind is broad, they cannot explain a large fraction of XRFs, and must rely on extrinsic causes, such as viewing-angle effects.

Rees and Mészáros (2005) considered dissipative effects (magnetic reconnection, neutron decay, or shocks) occurring in the optically thick inner parts of the relativistic outflows producing gamma-ray bursts and X-ray flashes. They particularly empha-
size the comptonization of the thermal radiation flux that is advected from the base of the outflow. Such dissipative effects boost the energy density of the thermal radiation. The dissipation can lead to pair production, in which case the pairs create an effective photosphere farther out than the usual baryonic one. In a slow dissipation scenario, pair creation can be suppressed, and the effects are most important when dissipation occurs below the baryonic photosphere. In both cases an increased photospheric luminosity is obtained. They suggest that the spectral peak in gamma-ray bursts is essentially due to the comptonized thermal component from the photosphere, where the co-moving optical depth in the outflow falls to unity. Typical peak photon energies range between those of classical bursts and X-ray flashes. The relationship between the observed photon peak energy and the luminosity depends on the details of the dissipation, but under plausible assumptions can resemble the observed $E_{\text{peak}} - L_{\text{iso}}$ correlation.

The possible existence of intrinsic XRFs has been reinforced by the discovery of X-ray \textit{flares} following classical GRBs (see Section 4.9.2; these X-ray \textit{flares} are distinct from the X-ray flashes discussed in this section, the terminology is unfortunately very confusing). The X-ray \textit{flares}, whose properties are similar to those of X-ray flashes (see for instance Arimoto et al. 2007), have been attributed to continuing activity of the central engine. This suggests that the central engine of GRBs is fully capable of producing intrinsically soft bursts like XRFs.

In 2007, 6 years after their discovery, the nature of the link connecting XRFs with classical GRBs remains mysterious. Depending on the model considered, it is really possible that XRFs constitute the bulk of a population of high-energy transients. This is the result obtained by Pêlangeon et al. (2008) who computed the volume of detectability of 82 GRBs detected by the WXM and FREGATE on HETE-2, and inferred a spatial density of XRFs (12 Gpc$^{-3}$ yr$^{-1}$) which is significantly higher than the spatial density of classical GRBs (1 Gpc$^{-3}$ yr$^{-1}$). This result should be put in perspective with the existence of low-luminosity GRBs, of the type of GRB 980425 and GRB 060218 (see Section 4.10.1), which are still closer and fainter than XRFs ($z = 0.0085$ and $z = 0.033$, respectively). The spatial density of these low-luminosity GRBs is of the order of few 10$^2$ Gpc$^{-3}$ yr$^{-1}$ (Soderberg et al. 2006b, Guetta & Della Valle 2007). There is a growing feeling that the emission of relativistic jets by dying stars and the subsequent emission of bursts of high-energy photons is a very general mechanism. Thirty years after the discovery of GRBs, we are just starting to uncover the diverse appearance of this mechanism as classical GRBs, XRFs, and low-luminosity GRBs. In this framework, classical GRBs are just ‘monster events’ which were discovered first, due to their exceptional energy content. This is analogous with other astronomical sources of relativistic jets, the quasars and active galactic nuclei (AGN). Bright quasars were discovered early on, in the 1960s, while the bulk of the AGN population, and the universality of the phenomenon, were discovered many years later.

4.6 INTEGRAL AND THE X-RAY HALOS OF GRBS

INTEGRAL is a European multi-purpose gamma-ray mission, dedicated to the fine spectroscopy and fine imaging of celestial gamma-ray sources in the energy range
15 keV to 10 MeV, with concurrent source monitoring in the X-ray and optical energy ranges (Winkler et al. 2003). INTEGRAL carries two main gamma-ray instruments: SPI, a spectrometer optimized for high-resolution gamma-ray line spectroscopy in the energy range 20 keV to 8 MeV (Vedrenne et al. 2003), and IBIS, an imager optimized for high-angular resolution imaging in the range 15 keV to 10 MeV (Ubertini et al. 2003). The SPI anticoincidence system is also a very sensitive burst monitor—but with no spectral or localization capability (Rau et al. 2005). It is used mainly in the IPN. Two monitors, JEM-X (Lund et al. 2003) in the (3–35) keV X-ray band, and OMC (Mas-Hesse et al. 2003) in optical Johnson V-band complement the payload. INTEGRAL was launched on October 17, 2002. Even though it was not a GRB-dedicated mission, INTEGRAL-IBIS had a sufficient field of view to detect about one GRB per month, and the fast alert capability needed to allow follow-up from the ground. Unlike HETE-2 and Swift which were in low-Earth orbits, INTEGRAL is in a highly elliptical orbit allowing an almost continuous monitoring from the ground. This characteristic of the mission led Mereghetti et al. (2003) to set up a GRB alert process, called IBAS, based on the processing on the ground of images of the IBIS field of view transmitted by the spacecraft in near-real time (Mereghetti et al. 2003). IBAS permitted the distribution of GRB positions with an accuracy of a few arcminutes within about 10 s of the trigger. Like HETE-2 and Swift, INTEGRAL is detecting GRBs of all types: classical GRBs, short GRBs, X-ray flashes, dark bursts, etc. Unfortunately, many of these events did not have a good optical follow-up because INTEGRAL is often looking at the galactic disk, where the high galactic extinction prevents the detection of optical afterglows. On the other hand, INTEGRAL is on the same type of satellite platform as the powerful XMM X-ray telescope, and in a similar orbit, allowing XMM to follow up many INTEGRAL GRBs at X-ray wavelengths. We discuss below some original advances which have been made possible by INTEGRAL and its powerful combination with XMM.

Polarization of the prompt emission

As explained in Chapter 6, measuring the linear polarization of the prompt GRB emission would provide a powerful diagnostic of the emission mechanism and of the geometry of the jet. Such measurements can be performed on the prompt gamma-ray signal with instruments made of multiple detectors in a plane. This is the case with RHESSI with its nine germanium detectors (Lin et al. 2002), with SPI, the spectrometer of INTEGRAL with 19 germanium detectors (Vedrenne et al. 2003), and with PIsCIT with 4096 cesium iodine scintillator detectors (Ubertini et al. 2003). After the controversial claim of the detection of a nearly 100% linear polarization of the prompt emission of GRB 021206 with RHESSI (Coburn & Boggs 2003, Rutledge & Fox 2004, Wigger et al. 2004), it was important to try to measure the polarization of the prompt emission of another GRB. This has been possible with the detection of GRB 040219A with INTEGRAL-SPI. GRB 040219A is a long, bright GRB detected up to a few megaelectronvolts (McBreen et al. 2006). McGlynn et al. (2007) measured the polarization over several energy ranges and time intervals and found a signal consistent with a polarization of about 60%, but at a low level of significance (~2σ). Kalemc et al. (2007) did a similar analysis which yielded a polarization fraction of 98% ± 33%. They concluded that, statistically, they could not claim a polarization
detection from this source. Also, they could not strongly rule out the possibility that the measured modulation was dominated by instrumental systematic effects. The conclusion of these two studies was that SPI observations of GRB 041219A could not significantly constrain the polarization of the prompt emission. Nevertheless, both authors concluded positively on the capability of SPI to measure polarization in future very bright GRBs. The polarization of prompt GRB emission was also studied with BATSE using the GRB photons Compton-scattered off the Earth’s atmosphere. Two bursts showed evidence of polarization, but the many systematic effects involved in the analysis did not allow definite conclusions to be reached (Willis et al. 2005).

X-ray lines in the afterglows?

Watson et al. (2003) have reported the detection of strong delayed X-ray line emission in the X-ray afterglow of GRB 030227, 20 h after the burst. The delayed arrival of the lines requires some particular geometry/environment of the burster and/or late energy injection by the central engine. This delayed arrival may also explain the lack of reported line detections with the XRT of Swift. The real level of significance of these lines is, however, still a matter of debate (see Butler et al. 2005b and Sako, Harrison, & Rutledge 2005), and more data are clearly needed to conclude securely on the existence and nature of X-ray lines in GRB afterglows.

GRB 031203, a nearby GRB with a hypernova

GRB 031203 was detected as a 20s long GRB with INTEGRAL and quickly localized (Gotz et al. 2003, Sazonov, Lutovinov, & Sunyaev 2004). An XMM observation started 6 h after the burst revealed two sources within the error box (Santos Lleo, Calderon, & Gotz 2003). The source closer to the INTEGRAL position coincided with a galaxy, which was later found to be fading and associated with a radio transient (Frail 2003, Soderberg, Kulkarni, & Frail 2003). While no optical afterglow was detected, the spatial coincidence of the X-ray afterglow and the radio afterglow with a nearby galaxy left no doubt about the identification of the host galaxy. The spectroscopy of the host showed it to be nearby at redshift $z = 0.105$ (Prochaska et al. 2003). The host is an actively star-forming galaxy, typical of GRB hosts (Prochaska et al. 2004). At the time of its discovery GRB 031203 aroused great interest because it was the second-nearest GRB after GRB 980425. This low redshift prompted searches for a supernova accompanying the GRB, and the searches were indeed successful, revealing a bright hypernova, spectroscopically and photometrically analogous to 1998bw, but brighter (Cobb et al. 2004, Gal-Yam et al. 2004, Malesani et al. 2004, Thomsen et al. 2004, Mazzali et al. 2006b). This supernova was called 2003lw.

While GRB 031203 was closer than GRB 030329, it was about 50 times fainter, with a fluence of $2 \times 10^{-6}$ erg cm$^{-2}$ (20–200 keV), compared with the fluence $1.2 \times 10^{-4}$ (30–400 keV) of GRB 030329. In terms of energy, GRB 031203 had an isotropic equivalent energy $E_{\text{iso}} \sim 10^{50}$ erg in gamma-rays, clearly less than the bulk of the typical GRB population at $z \sim 1$. This detection of a nearby ‘under-luminous GRB’ after GRB 980425, raised interesting questions about the nature and spatial density of GRB sources. From the properties of the prompt high-energy emission and
the radio afterglow, Sazonov, Lutovinov, and Sunyaev (2004) and Soderberg, Kulkarni, and Frail (2003) inferred that GRB 031203 was an analogue of GRB 980425, and not a GRB viewed off-axis. They both concluded that a large population of such under-luminous GRBs should exist, which are only marginally detected by current instrumentation. By contrast, Watson et al. (2004, 2006c) claimed that the bright echo detected with XMM (see below) was due to the prompt GRB (not the early afterglow), implying a burst much softer than that inferred from the INTEGRAL-IBIS data. In this case GRB031203 would have a fluence ratio $S_x/S_\gamma = 1.8 \pm 0.4/0.9$, implying that it is an X-ray flash. In this case GRB 031203 would be ‘just another XRF’ whose properties can be compared with the canonical XRF 020903, which was also associated with a bright supernova in a nearby galaxy. The two bursts appear quite similar, with GRB 031203 having a redshift 2.5 times smaller, an equivalent energy $E_{iso}$ five times larger, and an $E_{peak}$ two–three times smaller (Watson et al. 2004). Unfortunately, the observational evidence was insufficient to resolve this issue.

**X-ray halos**

The localization of several INTEGRAL GRBs in the galactic plane, and the excellent synergy between INTEGRAL and XMM had one interesting consequence: the discovery of bright, time-dependent, dust-scattered X-ray halos around GRBs. These halos are due to the scattering of X-rays by dust in molecular clouds in our own Galaxy. The first halo was observed around GRB 031203 (Vaughan et al. 2004), which was observed by XMM-Newton starting 6 h after the burst. The halo appeared as concentric ringlike structures centered on the GRB location (see Figure 4.7). The radii of these structures increased with time as $t^{1/2}$, consistent with small-angle X-ray scattering caused by a large column of dust along the line of sight to a cosmologically distant GRB. These rings are due to dust concentrated in two distinct regions in the Galaxy, located at distances of 880 and 1390 pc, consistent with known Galactic features. Tiengo and Mereghetti (2006) proposed an efficient way to find faint halos, and reported the discovery of another halo around GRB 050713A. In this case the dust layer is located at a distance of 364 pc. After these pioneering discoveries, more halos have been discovered in the Swift era: around the short GRB 050724 (Romano et al. 2005, Vaughan et al. 2006b), GRB 061019, and GRB 070129 (Vianello, Tiengo, & Mereghetti 2007). The detection of such halos offers a powerful tool for mapping the three-dimensional distribution of Galactic dust clouds. When the prompt X-ray emission (the one which is scattered) and the halo are observed the method has the potential to measure some properties of the scattering interstellar dust (Vianello, Tiengo, & Mereghetti 2007).

### 4.7 THE SWIFT MISSION

The Swift satellite was launched on November 20, 2004 and became fully operational in April 2005. It has been designed for a multi-wavelength approach of GRB observations. It includes three instruments: BAT, the Burst Alert Telescope, XRT, the X-Ray Telescope and UVOT, the Ultra-Violet/Optical Telescope. Figure 4.8 shows a drawing of the Swift Satellite. Swift has the unique capability of performing
Figure 4.7. X-ray images of GRB 031203 observed with Epic on XMM-Newton and covering the 0.7–2.5 keV energy range. These observations span a period of time extending from 6 h to 16 h after the burst. The X-ray halo appears as concentric ring-like structures centered on the GRB position (Vaughan et al. 2004). See also Color section.

Figure 4.8. Schematic view of the Swift satellite (Gehrels et al. 2004). The size of the coded mask of BAT is 2.7 m². See also Color section.
fast autonomous slew to track the GRB and its afterglow with minimum delay. The X-ray counterpart can be observed with the XRT as early as 50–70 s after the trigger and followed for days to weeks. For UVOT, the delay in observing the GRB is a little longer (80–100 s typically). A complete description of the instruments, with the main references, is given in Gehrels et al. (2004). We report below the main characteristics of the three instruments and reproduce Tables 2, 3, and 4 of Gehrels et al. (2004) that summarize their characteristics.

### 4.7.1 BAT: the Burst Alert Telescope

The detector is a mosaic of CdZnTe (CZT) detectors, made of 32 768 elements of $4 \times 4 \times 2$ mm, offering a sensitive area of 5240 cm$^2$ (Barthelmy 2000, 2004; Table 4.1). This configuration is similar to that of the CdTe detectors chosen for IBIS on INTEGRAL but the area is increased by a factor of $\sim$2. This detector is completed by a D-shaped coded mask made of 54 000 lead tiles ($5 \times 5 \times 1$ mm) mounted on a 5 cm thick composite honeycomb panel. The mask is located 1 m above the detector plane. The pattern of the mask is based on 50% open and 50% closed elements. The area of the mask is 2.7 m$^2$, yielding a half-coded field of view of $100^\circ \times 60^\circ$, or 1.4 sr.

<table>
<thead>
<tr>
<th>BAT parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy range</td>
<td>15–150 keV</td>
</tr>
<tr>
<td>Energy resolution</td>
<td>$\sim$7 keV</td>
</tr>
<tr>
<td>Aperture</td>
<td>Coded mask, random pattern, 50% open</td>
</tr>
<tr>
<td>Detection area</td>
<td>5240 cm$^2$</td>
</tr>
<tr>
<td>Detector material</td>
<td>CdZnTe (CZT)</td>
</tr>
<tr>
<td>Detector operation</td>
<td>Photon counting</td>
</tr>
<tr>
<td>Field of view</td>
<td>1.4 sr (half-coded)</td>
</tr>
<tr>
<td>Detector elements</td>
<td>256 modules of 128 elements module$^{-1}$</td>
</tr>
<tr>
<td>Detector element size</td>
<td>$4 \times 4 \times 2$ mm$^3$</td>
</tr>
<tr>
<td>Coded-mask cell size</td>
<td>$5 \times 5 \times 1$ mm$^3$ Pb tiles</td>
</tr>
<tr>
<td>Telescope PSF</td>
<td>$&lt;20''$</td>
</tr>
<tr>
<td>Source position and determination</td>
<td>$1'–4'$</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$\sim$10$^{-8}$ erg cm$^{-2}$ s$^{-1}$</td>
</tr>
<tr>
<td>Number of bursts detected</td>
<td>$&gt;100$ yr$^{-1}$</td>
</tr>
</tbody>
</table>
A graded-Z fringe shield located under the detector and surrounding the mask and the detector plane (Figure 4.9) is used to reduce the isotropic cosmic diffuse X-ray flux and the contribution of the earth-albedo, the orbit being low, 600 km with an inclination of 22°. The burst trigger algorithm looks for excesses in the detector count rate above the expected background (diffuse background + constant X-ray sources). A key feature of the BAT instrument is its imaging capability. The on-board software requires that the trigger corresponds to a point source, eliminating many spurious events, which could lead to a significant increase of the background of GRB-like events. This imaging capability permits an autonomous trigger of the spacecraft and its slew to the GRB position typically determined with a precision of few arcminutes. This automatic and rapid slew (50° in less than 75 s) allows pointing the XRT and UVOT at the GRB position with a minimum delay (considerably less than for BeppoSAX which was not designed for that). The energy range of BAT is 15–150 keV (the non-coded response extends up to 500 keV) and the energy resolution is 3.3 keV at 60 keV ($\Delta E / E \sim 5\%$). The sensitivity is about 2.5 times better than BATSE. The angular resolution is 20′ FWHM, with a 1–3′ centroiding capability. When a burst is detected, its celestial coordinates and intensity are immediately sent to the ground and distributed to the scientific community through the GRB coordinates network (GCN).

### 4.7.2 XRT: the X-Ray Telescope

XRT is a focusing X-ray telescope with a 110 cm$^2$ effective area, 23′ field of view and 18′′ resolution (Table 4.2). The energy range extends from 0.2 to 10 keV. This instrument uses a grazing incidence Wolter I telescope to focus soft X-rays onto a CCD detector. The X-ray mirrors are the flight spares for the JET X instrument on the Spectrum X mission (Wells et al. 1992, 1997). The CCD detectors were designed for the EPIC-MOS instrument on the XMM-Newton mission. The high fluxes of GRB afterglows permit the determination of source positions with a precision of 5-6′′ in flight and 2–3′′ using ground analysis. The sensitivity and imaging capabilities are respectively five and four times better than BeppoSAX narrow-field instruments. The
Table 4.2. Main characteristics of XRT on Swift.

<table>
<thead>
<tr>
<th>XRT parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy range</td>
<td>0.2–10 keV</td>
</tr>
<tr>
<td>Telescope</td>
<td>JET-X Wolter 1</td>
</tr>
<tr>
<td>Detector</td>
<td>E2V CCD-22</td>
</tr>
<tr>
<td>Effective area</td>
<td>110 cm$^2$ at 1.5 keV</td>
</tr>
<tr>
<td>Detector operation</td>
<td>Photon counting, integrated imaging, and timing</td>
</tr>
<tr>
<td>Field of view</td>
<td>$23.6' \times 23.6'$</td>
</tr>
<tr>
<td>Detection elements</td>
<td>$600 \times 602$ pixels</td>
</tr>
<tr>
<td>Pixel scale</td>
<td>2.26&quot;</td>
</tr>
<tr>
<td>Telescope PSF</td>
<td>18&quot; HPD at 1.5 keV</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$2 \times 10^{-14}$ erg cm$^{-2}$ s$^{-1}$ (1 mcrab) in $10^4$ s</td>
</tr>
</tbody>
</table>

The major advantage of Swift is certainly its fast slewing capability which allows pointing the XRT at BAT GRB positions in typically less than 100 s after the burst trigger.

4.7.3 UVOT: the Ultra-Violet/Optical Telescope

UVOT is based on the optical monitor used on-board ESA’s XMM-Newton mission (Mason et al. 1996, 2001). It is co-aligned with the XRT, allowing broadband UV-visible photometry and low-resolution grism spectra of bright GRBs in a range of wavelengths extending from 170 to 600 nm. UVOT is a 30-cm clear aperture Ritchey Chretien telescope. Two redundant photon-counting microchannel-intensified CCD detectors are selected by a steerable mirror mechanism. Each detector has a filter wheel, with 11 positions. Table 4.3 gives the main characteristics of UVOT, among them the field of view $17' \times 17'$ and sub-arcsec positioning allowing accurate determination of the position of a GRB relative to its host galaxy. In flight the sensitivity is typically 22nd magnitude.

4.7.4 The alert system

Swift GRB alerts and burst properties are transmitted to the ground almost instantaneously through a TDRSS link. Their rapid distribution to the international community is then ensured by the GCN. A TDRSS uplink also allows rapid response to TOOs such as GRBs detected by other missions.

Swift has demonstrated its ability to analyze GRB afterglows with an unprecedented rapidity (less than 100 s after the identification of the burst by BAT)
Table 4.3. Main characteristics of UVOT on Swift.

<table>
<thead>
<tr>
<th>UVOT parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength range</td>
<td>170–600 nm</td>
</tr>
<tr>
<td>Telescope</td>
<td>Modified Ritchey–Chrétien</td>
</tr>
<tr>
<td>Aperture</td>
<td>30 cm diameter</td>
</tr>
<tr>
<td>f' number</td>
<td>12.7</td>
</tr>
<tr>
<td>Detector</td>
<td>Intensified CCD</td>
</tr>
<tr>
<td>Detector operation</td>
<td>Photon counting</td>
</tr>
<tr>
<td>Field of view</td>
<td>17' x 17'</td>
</tr>
<tr>
<td>Detection elements</td>
<td>2048 x 2048 pixels</td>
</tr>
<tr>
<td>Telescope PSF</td>
<td>0''9 FWHM at 350 nm</td>
</tr>
<tr>
<td>Colors</td>
<td>6</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>$B = 24$ in white light in 1000 s</td>
</tr>
<tr>
<td>Pixel scale</td>
<td>0.5''</td>
</tr>
</tbody>
</table>

to discover the properties of the early afterglows (see the following sections) and to follow the X-ray afterglows over durations never reached with BeppoSAX. This is very important for studying the X-ray light-curves in search of the jet breaks found at optical wavelengths. With the large effective area of BAT (5240 cm$^2$), Swift is better suited than previous missions for finding bursts with low peak flux or fluence. Let us go now to a short review of the results of this mission, the second to be completely designed for GRB studies, after HETE-2.

4.8 A GENERAL ANALYSIS OF THE FIRST GRBS OF SWIFT

Swift localizes ~90 GRBs per year, or about two GRBs per week, to be compared with the HETE-2 and BeppoSAX rates of about two GRBs per month. This rate is compatible with the fact that the sensitivity of Swift is slightly better than that of BATSE.

Swift afterglows are fainter than those of BeppoSAX or HETE-2. The average fluxes of optical and X-ray afterglows at $t = 12$ h are: $R \sim 21.5$ mag and $F_x \sim 3 \times 10^{-13}$ erg cm$^{-2}$ s$^{-1}$ (Berger et al. 2005). These values are respectively 1.7 magnitudes (a factor of 4.7) and a factor of 3 fainter than pre-Swift afterglows. The higher sensitivity and accurate positions of Swift result in a more accurate representation of the true burst redshift and brightness distributions, which are higher and dimmer, respectively (Berger et al. 2005a).
Figure 4.10. Cumulative distribution of GRBs as a function of redshift for 43 pre-Swift bursts (upper stepwise curve) and for 19 Swift bursts (lower stepwise curve), from Jakobsson et al. (2006b). The observed curves are compared with theoretical models: two models from Natarajan et al. (2005) which assume a GRB rate proportional to the star-formation rate (solid curve) and a GRB rate increasing with decreasing metallicity (dashed curve), and one model from Gorosabel et al. (2004) assuming a GRB rate proportional to the star-formation rate (dash-dotted curve).

Another point is the absence of a large sample of nearby low luminosity events \((z < 0.2)\). So a local population of low-energy bursts does not contribute significantly to the sample. On the other hand, Swift observes a population of high-redshift bursts which is larger than that of BeppoSAX or HETE-2, as we discuss now.

The Swift GRB population peaks at a higher redshift than pre-Swift bursts. Figure 4.10 gives the cumulative fraction of GRBs as a function of redshift for 43 pre-Swift bursts and 19 Swift bursts. The median redshift of Swift GRBs is \(z \sim 2.72\), more than two times larger than the corresponding value for pre-Swift bursts \((z \sim 1.04)\). This is explained by the lower trigger threshold of Swift and therefore its ability to detect fainter bursts. Even if a priori it was not certain that fainter long bursts are on average at larger redshifts, Bagoly et al. (2006) have shown that this is indeed the case. Using BATSE data many authors have suggested that large redshifts (up to \(z = 20\)) could be expected for faint GRBs (see for instance Lamb & Reichart 2000, Bagoly et al. 2003, Lin, Zhang, & Li 2004). Moreover the accurate positions given by the Swift XRT, associated with the rapid response of a large variety of telescopes aimed at redshift determinations, might have contributed to this higher average redshift.

Among the general considerations that can be made when comparing Swift GRBs with the pre-Swift population, Berger et al. (2005a) noted the following points.
The overall detection fraction of X-ray afterglows for Swift bursts is 21/22; this is essentially the same value as that obtained for past missions (>90%).

The fraction of optical/near-IR afterglows for the Swift sample (11/21, ~52%) is higher than the 30% obtained by past missions but is worse than the 85% fraction obtained for the SXC bursts of HETE-2. Conversely, the fraction of dark bursts estimated by Berger et al. (2005a) is at least 10% and perhaps as high as 1/3. This could be the result of high redshift or dust extinction.

The radio follow-up of 13 Swift bursts led to the detection of only three radio afterglows. This is comparable to the fraction of previous missions (30%) and again lower than the 55% recovery of the SXC sample (Berger et al. 2005a).

Another property, which had already been noticed with the development of several robotic telescopes but was clearly demonstrated with Swift, is the paucity of very bright early optical afterglows. In fact, as noted by Roming et al. (2006a), early observations with UVOT revealed that in a large number of cases, the early optical afterglow is not detected. So, bright optical flashes like the one observed in GRB 990123 (see Section 3.9) are only present in a few GRBs. The possible mechanisms for this lack of early optical afterglows are analyzed by Roming et al. (2005); their conclusions are the following: for ~25% of the bursts in their sample the lack of optical emission can be attributed to dust in our Galaxy, for ~25% the absorption takes place in the by circumburst environment. An additional 30% are most likely attributable to Lyα blanketing and absorption at high redshift. Poynting flux dominated outflows or pure non-relativistic hydrodynamic reverse shocks could also suppress the early optical afterglow. Moreover, as we will see in Chapter 6, the physics of the reverse shock can in many cases lead to weaker optical emission than previously thought. The scarcity of these optical flashes might also be due to low values of $\nu_m$, the typical synchrotron frequency, which could originate from rather small values of the microphysical parameters $\varepsilon_e$ and $\varepsilon_B$ (Mundell et al. 2007). On the other hand, Roming et al. (2006a) rule out other mechanisms such as a rapid temporal decay or a low-density environment. In conclusion, the paucity of detected optical flashes is a real effect measured with UVOT and not a consequence of the observing conditions. To be fair, the scarcity of bright optical flashes was not a real surprise after many ground-based robotic telescope observations.

Let us go now to the analysis of the early X-ray light-curves which were terra incognita in the BeppoSAX era. Their behavior and the presence of many X-ray flares represent a major breakthrough of the Swift mission.

4.9 THE DIFFERENT PHASES OF THE X-RAY AFTERFLOWS

One major advance enabled by Swift is the discovery of the unexpected shape of X-ray early afterglows. This has been made possible thanks to the association of two major instruments on a satellite which was able to quickly calculate the GRB position and slew to it. BAT (Barthelmy et al. 2005c) was able to detect bursts with 15–150 keV fluences as low as $\sim10^{-8}$ erg cm$^{-2}$, and the X-ray Telescope (XRT; Burrows et al.
was able to follow them down the fluxes $\sim 10^{-13}$ erg cm$^{-2}$ in its energy range of 0.3–10 keV. In the pre-Swift epoch several hours were needed to observe the X-ray afterglows. Swift has the capability to slew much more rapidly, and to start the observations in less than 1 min after the $\gamma$-ray trigger. With Swift the coverage of the X-ray afterglow is dramatically increased with better sensitivity, resulting in a completely new view of the early X-ray afterglow (see Section 4.9.1). The parallel observations with the Ultra-Violet Optical Telescope (UVOT; Roming et al. 2005) complete this early multi-wavelength coverage.

Swift also provides the capability to observe the transition between the prompt emission (observed with BAT) and the early X-ray afterglow (observed with the XRT). Since BAT and XRT data cover different energy bands and seldom overlap in time, the comparison requires extrapolating the data of one instrument to the band pass of the other. To study this transition, O’Brien et al. (2006a) have chosen to use the 0.3–10 keV band pass, extrapolating BAT data to low energies (below 15 keV). They measure $\beta_p$, the BAT spectral index, and $\beta_x$, the early XRT spectral index. They show (Figure 1 of their paper) that there is a trend such that the early XRT data have systematically softer power-law indices ($\beta_p > \beta_x$). This is not surprising because of the well known trend of the GRB prompt $\gamma$-ray emission to soften with time. O’Brien et al. (2006a) have combined BAT and XRT data to study the unabsorbed 0.3–10 keV light-curves at the time of the transition between the prompt emission and the early afterglow (Figure 4.11). Their work (based on 40 GRBs) strongly suggests that there is no discontinuity between the emission seen by BAT and the XRT. In their sample of 40 GRBs they find only two exceptions GRB 050219A (where the discrepancy can be explained by an X-ray flare occurring during the slew), and GRB 050525A.

The second important point was the unanticipated behavior of the early X-ray afterglow. After the launch of Swift it appeared quickly that the X-ray light-curves of many GRBs displayed a steep early decay up to few hundred seconds followed by a period of flatter decay (Figure 4.12; Vaughan et al. 2006a, Barthelmy et al. 2005a, Cusumano et al. 2006b, Tagliaferri et al. 2005b). This behavior, although frequent, is not universal: see for instance Campana et al. (2005) who reported for GRB 050128 an early decay relatively flat and a steepening at later time ($\sim 1500$ s). Several studies made it possible to extract the general properties of these X-ray early afterglows (Chincarini et al. 2005, Nousek et al. 2006, O’Brien et al. 2006a, Panaitescu 2006a, Willingale et al. 2007). Let us discuss these properties.

### 4.9.1 The canonical early X-ray afterglow light-curve

After the observation of a few dozen X-ray afterglows with the XRT, a canonical X-ray light-curve appeared. This canonical evolution is illustrated in Figure 4.13(a) (from Zhang et al. 2006), it is found in about two-thirds of the GRBs (O’Brien et al. 2006b). In this figure parts I, II, and V (X-ray flares) are the new parts of the early afterglow revealed by Swift, while parts III and IV were already identified in the pre-Swift period (see also Figure 4.13(b), from O’Brien et al. 2006b illustrating two types of X-ray light-curves).
Figure 4.11. Examples of BAT and XRT light-curves for some Swift GRBs among 40 bursts analyzed by O'Brien et al. (2006a). The early X-ray emission is plotted out to $10^5$ s. Left-hand panels show the combined BAT + XRT unabsorbed 0.3–10 keV light-curves. Right-hand panels show the spectra relative to the power law derived from fitting the BAT data (O'Brien et al. 2006a).
In the early X-ray afterglow light-curve, one can identify three power-law segments \( F_v \propto v^{-\beta} t^{-\alpha} \) including a fast initial decay \( (3 < \alpha_1 < 5) \), a very shallow decay \( (0.5 < \alpha_2 < 1.0) \), and a phase with an intermediate decay \( (1 < \alpha_3 < 1.5) \). Between these power-law phases there are two breaks: \( t_{b,1} \) after the steep decay, and \( t_{b,2} \) after the shallow phase. Nousek et al. (2006) give the distribution of the spectral index \( \beta_x \) and of the temporal indices \( \alpha_1, \alpha_2, \) and \( \alpha_3 \) for 27 GRBs. A large dispersion of \( \alpha_1 \), the slope of the steep decay, is quite apparent with \( 3 < \alpha_1 < 5 \). On the other hand, the other exponents are much less dispersed: \( 0.5 < \alpha_2 < 1 \) and \( 1 < \alpha_3 < 1.5 \). \( \beta_x \) is mainly distributed between 0 and 1. Nousek et al. (2006) also give the distribution of \( t_{b,1} \) and \( t_{b,2} \): \( t_{b,1} \) is \(<500 \text{s} \) while \( t_{b,2} \) is largely dispersed with \( 10^3 \text{s} < t_{b,2} < 10^4 \text{s} \).

As we will see in Chapter 7, these early X-ray emissions can be explained; the steep decay phase would be due to the end of the emission of the central engine with the diminishing contribution of high-latitude emission (the curvature effect). The steep decay slope can be affected by an early afterglow contribution with its tendency to decrease the slope of the decay. The following shallow phase seems to be more naturally explained by a continued central engine activity (Nousek et al. 2006, Zhang et al. 2006) but other origins are possible, as discussed in Chapter 7. In fact a long-lasting central engine activity may also explain another characteristic of the early X-ray afterglow, the presence of X-ray flares in at least half of the GRBs, which are usually observed in the early X-ray afterglow phase. Let us turn now to this new component of the X-ray light curve.

4.9.2 X-ray flares

Before Swift, flaring activity had been observed in X-rays from tens of seconds to several minutes in GRB970228 (Frontera et al. 1998; see Chapter 3) and in GRB
Figure 4.13. (a) A cartoon X-ray light-curve proposed by Zhang et al. (2006) based on the observational data from the Swift XRT. Phase 0 denotes the prompt emission. I is the prompt decay phase, II the really new shallow phase with V indicating possible flares. Phase III was already observed by BeppoSAX with a possible break leading to phase IV. Segments I and III (solid lines) are the most common (from Zhang et al. 2006). (b) Another schematic view of the early X-ray light-curve with the two types of afterglows, proposed by O’Brien et al. (2006b).

011211 and GRB 011211, with late X-ray bursts at 240–310 s and 600–700 s after the trigger (Piro et al. 2005). XRF 011030 showed a long bursting activity lasting about 1500 s, with a late X-ray flare at ~1300 s (Galli & Piro 2006; see their Figures 1 and 2). So the flaring activity was identified before Swift, but what has been quite surprising with Swift was the presence of flares in a large percentage of X-ray light-curves, about
half of them (O'Brien et al. 2006b). Since many of these flares are seen in the early X-ray afterglow it is not surprising that they were not easily observed by BeppoSAX (those reported had been seen thanks to the WFC).

Burrows et al. (2005a) were the first to report two GRBs detected by Swift showing strong X-ray flares. The first of these was XRF 050406, an X-ray flash with a short and relatively weak X-ray flare which appeared 213 s after the prompt emission (see also Romano et al. 2006). When the underlying power-law decay is subtracted, the rise and fall of the flare are nearly symmetric with very steep temporal power-law indices of $\pm 6.8$, reminiscent of the steep decay phase we have reported for the early afterglow. The second case was reported for a typical GRB 050502B, which had an X-ray flare peaking 740 s after the prompt GRB emission. This is still the largest flare; its peak flux was $>500$ times that of the underlying afterglow (Figure 4.14, Falcone et al. 2006). The fluence of this X-ray flare in the 0.2–10 keV energy band exceeded the fluence of the prompt GRB and its spectrum was harder than those measured before and after the flare (Falcone et al. 2006). These authors suggested that this giant flare was the result of internal dissipation of energy due to a late central engine activity episode rather than an afterglow-related effect. These first two observations illustrate the fact that X-ray flares are a common feature of XRFs and GRBs, suggesting a potential link between these two classes of bursts (see Section 4.4). GRB050607 provided another example of flares with the presence of multiple X-ray flares (Pagani et al. 2006). In some GRBs multiple flares have been observed, with an underlying afterglow consistent with having the same slope before and after the flares. This was taken as an indication that the flares have nothing to do with the afterglow (see Figure 4.14).

These examples demonstrate that X-ray flares can be really energetic and have large flux variations. Flare fluences can be exceptionally high, up to 100% of the prompt fluence and the flare fluxes can vary on a very short timescale. Spectral analyses indicate that these flares are spectrally harder than the underlying continuum (Burrows et al. 2005a, Romano et al. 2006, Falcone et al. 2006) with a hard-to-soft evolution which is also observed for the prompt emission (Pagani et al. 2006).

Another interesting point is the presence of flares in short GRBs. GRB 050724, for instance, exhibited several flare-like features, in particular a broad bump with a peak at $\sim 5 \times 10^4$ s (Barthelmy et al. 2005b, Campana et al. 2006b).

GRB050904, a long GRB at $z \sim 6.29$, also presents X-ray flares with a significant flaring activity lasting as long as $\sim 5000$ s (586 s in the rest-frame; Cusumano et al. 2006a).

These flares, which seem a quite general property of GRBs, are analyzed and their properties discussed by Falcone et al. (2006). The characteristics of these flares, their temporal properties and morphology, were also analyzed by Chincarini et al. (2007) for a large GRB sample. Their very complete survey of X-ray flares gives a good idea of their properties (they have analyzed a total of 69 flares). Figure 4.15 provides some examples of the light-curves they have studied. This figure nicely illustrates the shape, intensity, and time of these flares, which can be unique or multiple in one GRB. The authors calculated an equivalent width for the flares,
The different phases of the X-ray afterflows

Figure 4.14. (a) Light-curve of the X-ray afterglow of GRB 050502B in the energy range 0.2–10 keV. The giant flare starting at 345 ± 30 s corresponds to a rate increase by a factor of >500. (b) Fit of the flare and the underlying decay. The decay before and after the flare do not deviate significantly from fitting all underlying data simultaneously (from Falcone et al. 2006).
Figure 4.15. Several examples of X-ray flares. They can be single or multiple in a given GRB (from Chincarini et al. 2007).
which represents the time needed for integration of the continuum to collect the same fluence as of the flare. This parameter gives an indication of the lowest fluence that can be measured for a flare. Their Figure 2 gives the distribution of the equivalent width for the sample. They conclude that the flare is generally stronger than the continuum light-curve and possibly an unrelated phenomenon. Adopting a multiply-broken power law to model the underlying continuum and a number of Gaussians to model the superimposed flares, they fitted the different afterglows and summarized the parameters of the continuum and the flares. These fits allowed them to calculate $\Delta t/t$ for each flare, where $\Delta t$ is the Gaussian width and $t$ the peak time. The distribution of $t$ ranges between 95 s and $\sim 75\ 000$ s. The distribution of $\Delta t/t$ has a mean value $\langle \Delta t/t \rangle = 0.13 \pm 0.10$ (Figure 4.16).

The temporal behavior of each flare is studied using the rising and decaying slopes $\alpha_1$ and $\alpha_2$ and also $\Delta t_{\text{rise}} = t_p - t_1$ and $\Delta t_{\text{fall}} = t_2 - t_p$ with $t_1$ and $t_2$ being the beginning and the end of the flare, and $t_p$ the time of the peak (see how the authors determined these times and how critical they are to describing the temporal behavior of the burst and its decay slope). The knowledge of $\Delta t_{\text{fall}}/\Delta t_{\text{rise}}$ is interesting because simulations have shown that bursts due to internal shocks present a ratio

![Figure 4.16](image)

**Figure 4.16.** Scatter plot of flare parameters. $\Delta F/F$ is the flux of the flare divided by the flux of the underlying afterglow. $\Delta t/t$ is the duration of the flare divided by the time of its maximum after the trigger (from Chincarini et al. 2007). They used the FWHM of Gaussian fits of the flares as $\Delta t$, and the Gaussian peak time for $t$. The ratio of the peak flux over the underlying continuum flux was calculated using the best-fit models. The lines show the kinematically allowed regions for afterglow variabilities according to Ioka, Kobayashi, and Zhang (2005). The four limits plotted are based on (a) eq. (4) in Ioka, Kobayashi, and Zhang (2005) for dips (shown on axis), (b) eq. (7) in Ioka, Kobayashi, and Zhang (2005) for bumps due to density fluctuations (on axis), (c) for bumps due to patchy shells, and (d) for bumps due to refreshed shocks. The off-axis cases (viewing angle larger than the half-angular size of the variable region) are shown by dashed lines.
\( \Delta t_{\text{fall}} / \Delta t_{\text{rise}} \sim 3.4 \) (Daigne & Mochkovitch 1998). Figure 7 of Chincarini et al. (2007) shows that the distribution of \( \Delta t_{\text{fall}} / \Delta t_{\text{rise}} \) is clustered around 2, below 3.4. The authors also try to investigate if there is a link between the properties of the pulses of the GRBs detected by BAT and the X-ray flares detected by the XRT. Of course a link between the two would allow establishment of a common origin for these two components. They did not find any correlation between the number of gamma-ray pulses and the number of X-ray flares. The most common case is when the burst presents one single pulse followed by one or two X-ray flares. Nothing can be inferred about the number of X-ray flares from the number of gamma-ray pulses and vice versa.

The authors searched for other correlations, among them:

(a) A correlation between the peak intensity and the peak time. Their Figure 12 shows that late flares have a peak intensity which is lower than early ones. Moreover, as we have seen, the distribution of \( \Delta t / t \) is peaked around 0.13, meaning that late flares have a larger \( \Delta t \) so their fluence can be larger.

(b) In their Figure 14, Chincarini et al. (2007) found a possible correlation between \( \Delta t_{\text{fall}} / \Delta t_{\text{rise}} \) already defined and \( T_{90} \). This correlation is interesting because it was observed by Daigne and Mochkovitch (1998) in their simulations of the prompt emission.

(c) A correlation between the quiescent time (between two successive pulses or between two flares) and the peak brightness of the following event—nothing is observed.

(d) A correlation between the quiescent time separating two events and the ratio between the peaks of two successive events (their Figure 10). Flares and \( \gamma \)-ray pulses have the same peak ratio and their Figure 11 shows that the distribution of the ratio between the peaks of two successive events for X-ray flares is quite consistent with that of \( \gamma \)-ray pulses. These similarities in the behaviors of the two phenomena point in favor of a common origin for \( \gamma \)-ray pulses and X-ray flares.

(e) Going in the same direction Liang et al. (2006) have shown that if it is assumed that the post-flare decay index \( (\alpha_{\text{fall}}) \) satisfies the curvature effect \( \alpha_{\text{fall}} = \beta + 2 \) (see Section 7.1.1), the required time \( T_0 \) is at the start of the corresponding X-ray flare for several flares. This is another argument to support the curvature effect and the internal origin of the flares.

In conclusion the flares appear as random events not directly related to the way the prompt emission develops in time. Most of the indications obtained through the statistical analysis of Chincarini et al. (2007) indicate that the flare activity is similar to that of the prompt emission and that for a significant fraction of the flares they cannot be related to external shock mechanisms confirming an earlier conclusion given by Chincarini et al. (2006) in a preliminary presentation of their data set. The models which have been proposed to explain the X-ray flares are presented in Chapter 7 but the conclusion of this work is that X-ray flares can be explained by long-lasting engine activity. Hence, two features of the X-ray afterglows discovered by Swift, the shallow phase and the X-ray flares, seem to be connected with a
long-lasting activity of the central engine. But we will see in Chapter 7 that other unconventional models have also been proposed recently that could explain their properties (Genet, Daigne, & Mochkovitch 2007, Uhm & Beloborodov 2007). The explanation of the early X-ray afterglow with its different components is really an open issue. More statistical studies such as those we have reported (Nousek et al. 2006, O’Brien et al. 2006a, Chincarini et al. 2007) are needed to resolve this important question.

4.9.3 Achromatic X-ray breaks?

A much-awaited result of Swift observations was the confirmation in X-rays of the jet breaks detected in the optical few days after the burst (Section 3.4). None of the breaks discussed so far (between the steep decay and the shallow decay phases or between the shallow decay phase and the more classical decay) could be identified with the jet breaks found in the optical. The jets breaks have to be searched for at later times. This search faced two problems which have raised a lot of confusion on this topic: after few days the X-ray afterglow may be too faint to allow the unambiguous detection of a break, and some GRBs with bright X-ray afterglows have shown no X-ray breaks for many days after the burst. The issues raised by these observations are discussed at length in Section 7.5.

Let us now go to the description of some individual GRBs whose properties are particularly interesting and can be considered as major discoveries of the Swift mission.

4.10 SOME EXCITING GRBS

4.10.1 GRB 060218, the first low-redshift X-ray flash

This burst, discovered on February 18, 2006, is presently the XRF with the lowest redshift at $z = 0.033$ (Pian et al. 2006, Campana et al. 2006a, Sollerman et al. 2006, Modjaz et al. 2006, Mirabal et al. 2006), and the GRB with the second lowest redshift, after GRB 980425 associated with SN 1998bw at $z = 0.0085$. This event is particularly interesting for many reasons.

(a) It provides the best example of a supernova, SN 2006aj, associated with an XRF. This association extends the GRB/SN connection to faint GRBs and faint SNe, implying a common origin (Pian et al. 2006, Sollerman et al. 2006). This association was not the first, as it came after many other associations previously reported, arguing for a common progenitor for GRBs and XRFs—XRF 020903, for instance, had a light curve and spectrum consistent with a SN at $z = 0.25$, the redshift of the GRB. The difference between SNe associated with GRBs and SNe associated with XRFs might be related to the mass of the exploding stars. For SN 2006aj a relatively low initial mass of $20M_\odot$, evolving into a $3.3M_\odot$ CO star, has been proposed by
Mazzali et al. (2006a). Such an initial mass is lower than those estimated for the typical GRB/SNe (Woosley & Heger, 2006).

The location of XRF 060218 within a low-metallicity dwarf galaxy (Modjaz et al. 2006, Mirabal et al. 2006) might indicate that different metallicities of the progenitors explain the variety in massive stellar explosions (Mirabal et al. 2006). Hence as proposed by Mirabal, Halpern, and O’Brien (2007), this XRF seems to paint a picture in which sub-luminous, sub-energetic GRB born in low metallicity dwarf galaxies dominate the local ($z < 0.5$) population of GRB events (see also Soderberg et al. 2004b). A complementary remark which can be made is that low-metallicity progenitors in the local Universe may differ from the progenitors of classical GRBs at higher redshifts.

Fynbo et al. (2004) and Bersier et al. (2006) found that XRF 020903 and XRF 030723 could be considered to be GRBs seen off the jet axis. The off-axis jet model has also been considered to be unusually faint GRBs, like GRB 031203 (Ramirez-Ruiz et al. 2005) and GRB 980425, the first and nearest GRB (36 Mpc) associated with a SN event (Yamazaki, Ioka, & Nakamura 2003). Cobb et al. (2006) found that for XRF 060218 this scenario is implausible at the 98% level. In addition, for this burst at $\sim$145 Mpc low values of $E_{\text{iso}} = (6.2 \pm 0.3) \times 10^{49}$ erg (three orders of magnitude smaller than that of a typical long duration GRB), and of the peak energy $E_p = 4.9^{+0.4}_{-0.3}$ keV have been measured (Campana et al. 2006a). These values are consistent with the Amati correlation, also suggesting that this burst is not an off-axis event (Amati et al. 2007). Pian et al. (2006) arrived at the same conclusion.

Fan, Piran, and Xu (2006) suggested that the X-ray afterglow of this burst could be attributed to a continued activity of the central engine, which could be due to fall back accretion within the collapsar model. These authors considered that such ‘central engine afterglow’ may be common in under-luminous GRBs where the kinetic energy of the blast wave is small and the external shock does not dominate over this component. Such under-luminous bursts might be common but rarely seen because they are detectable only from short distances (Fan, Piran, & Xu 2006). This is also the conclusion of Sazonov, Lutovinov, and Sunyaev (2004) who consider that there exists a large population of under-luminous events such as GRB 031203. For Pian et al. (2006) and Soderberg et al. (2006b), the properties of XRF 060218 also suggest the existence of a population of events less luminous than classical GRBs, but possibly much more numerous and with lower radio luminosities (see also Soderberg et al. 2004b). These events would be the most abundant form of X-ray or $\gamma$-ray explosive events in the Universe but instrumental limits allow their detection only locally. So, Pian et al. (2006) concluded that the fraction of supernovae associated with GRBs or XRFs might be higher than generally thought. They estimated a local GRB rate of $110^{+180}_{-20}$ Gpc$^{-3}$ yr$^{-1}$, compared with 1 Gpc$^{-3}$ yr$^{-1}$ estimated from cosmological events only (see also the work of Pêlangeon et al. (2008) who determine the true event rates of XRFs and GRBs in the Universe, Section 4.5.3).

(b) Early photometric and spectroscopic data on XRF 060218/SN 2006aj have clearly revealed that this SN is a fast-evolving, broad-lined, type Ic SN, with an extremely short rise time and a large optical peak luminosity (Modjaz et al. 2006).
Figure 4.17. Spectrum of GRB 060218/SN 2006aj taken with the 6.5-m Multiple Mirror Telescope (MMT) at $\Delta t = 12.85$ days after the burst and $\sim 3$ days after the maximum light. The host galaxy emission lines have been removed. The spectrum has been scaled and shifted. From the comparison it appears that the supernovae SN 1997ef and SN 2002ap are better matches than SN 1998bw (from Modjaz et al. 2006).

The large expansion velocities are nevertheless smaller than those found for SN 1998bw. The spectra obtained with the 6.5 m Multiple Mirror Telescope at Mt. Whipple Observatory at $\Delta t = 12.85$ days are compared with spectra of other type Ic SNe at similar phases (Figure 4.17; Modjaz et al. 2006). Optical spectroscopy allowed the earliest detection of weak, supernova-like Si II near 5720 $\AA$ (with a velocity $v \sim 0.1c$), starting only 1.95 days after the burst trigger (Mirabal et al. 2006). In fact, UBVRI photometry obtained between 1 and 26 days suggests a short time delay between the GRB and the onset of the supernova (Mirabal et al. 2006). Moreover, from the light curves and the spectral evolution it is clear that the emission is dominated by the SN rather than by the afterglow from the very beginning. Figure 4.18 (Campana et al. 2006a) shows the XRT light-curve with a large bump, which largely dominates the afterglow contribution. The behavior of the afterglow with a power-law decay with a typical index $-1.2 \pm 0.1$ appears only after $10^4$ s. This behavior is similar to SN 1998bw where no optical afterglow was ever detected, contrary to SN 2003dh (Sollerman et al. 2006). So, it appears that the supernova light-curve peaked earlier than most known supernovae and its time origin can be fixed within less than one day from the GRB trigger.

SN 2006aj is the fastest-evolving but one of the least luminous GRB–SN discovered so far (2006), only 70% as luminous as SN 1998bw. It can be placed at the
faint end of the GRB–SN luminosity distribution which, after extinction corrections, extends from 0.6 to 2 times the luminosity of SN 1998bw. In the context of the luminosities of local type Ic SNe without GRBs, SN 2006aj is still at the bright end of the luminosity distribution. Clearly this SN follows the general rule that GRB SNe tend to be more luminous than (local) type Ic SNe which give no GRB (Ferrero et al. 2006). In fact, this fast-evolving SN has given the first and unique opportunity to observe a GRB and its connected supernova in X-rays, UV and optical bands, starting only \( \sim 100 \) s after the GRB.

The XRT light-curve shows the presence of a thermal component (with a typical temperature of about 2 million degrees), which can be attributed to the radiation emitted by a shock-heated plasma (Campana et al. 2006a). The contribution of this emission increases with time and becomes dominant at the end of the exponential decay. This thermal component cools and shifts into UV/optical bands as time passes. This is illustrated in Figure 4.18 (lower panel) (Campana et al. 2006a), showing the UVOT light-curve, and in Figure 4.19 showing the evolution of the temperature and radius of the soft thermal component. The temperature starts to decrease around \( 10^4 \) s, while the radius suddenly increases. The early appearance of this soft component in the X-ray spectrum is understood as a ‘shock breakout’ (Mirabal et al. 2006, Campana et al. 2006a). In supernovae triggered by core collapse a shock wave is generated which propagates through the progenitor star and ejects

**Figure 4.18.** XRT and UVOT light-curves of GRB 060218 (from Campana et al. 2006a). Upper panel: The XRT light-curve (0.3–10 keV) is shown with black circles. The contribution of the black-body component is indicated with gray circles. It becomes dominant at the end of the exponential decay. At about 10 000 s the light-curve breaks to a shallower power-law decay with an index of \(-1.2 \pm 0.1\) typical of X-ray GRB afterglows. Lower panel: The light-curves obtained with UVOT for different filters. From \( 10^4 \) to \( 10^5 \) s the flux is mainly due to the shock breakout, while it is dominated by the SN light after \( 2 \times 10^5 \) s.
the envelope. A very bright UV/X-ray burst of radiation is expected at the emergence of the shock wave through the surface of the star. This outburst had never been observed before SN 2006aj, because the breakout occurs without warning (except from the neutrinos) and lasts a very short time. Nevertheless its effects on the surrounding circumstellar material can be studied. This has been done, for instance by Ensmann and Burrows (1992) for SN 1987A. Hence, through the very early study of this GRB-associated supernova, it has been possible to obtain for the first time a direct measurement of the breakout of the shock wave driven by a semi-relativistic shell into the dense wind surrounding the WR progenitor. This important question was investigated theoretically a long time ago, by Colgate (1968, 1974), and more recently by Tan, Matzner, and McKee (2001).

XRF 060218 was thus an important discovery of Swift. Not only has it provided a better understanding of the GRB–XRF connection, but above all it has represented a milestone for the study of the very early emission of type Ic supernovae. After this very nearby GRB, let us now go to another extreme, the detection of the most distant gamma-ray burst ever observed, GRB 050904.²

4.10.2 GRB 050904: the most distant GRB

This was a long-duration burst (duration 220 s) detected by BAT on September 4, 2005. Its localization, a 4' radius obtained 81 s after the trigger, was quickly disseminated via the GRB Coordinates Network (http://gcn.gsfc.nasa.gov/). Swift’s XRT automatically slewed to the BAT position and 76 min after the burst a better localization was obtained, with a 6'' radius, and distributed to observers. Haislip et al.

²This changed on September 13, 2008, with the detection of GRB 080913 by Swift and the subsequent measurement of its redshift, z = 6.7, with GROND (Greiner et al. 2008).
(2006) observed over the next few hours the XRT localization at both near-infrared (NIR) and visible wavelengths. In the NIR they discovered a bright ($J = 17.4$ mag at 3.1 h after the burst) and fading source using the 4.1-m Southern Observatory for Astrophysical Research (SOAR) telescope in Chile. On the other hand, at visible wavelengths they did not detect the afterglow down to relatively deep limiting magnitudes, implying either a very high redshift or a highly dusty environment. Haislip et al. (2006) ruled out the dust extinction hypothesis and identified this burst as a very high redshift GRB with $z$ between 6 and 8 (Haislip et al. 2005a, Reichart 2005). With shorter wavelength NIR detections with SOAR, they quickly narrowed this range to its lower end: $z \sim 6$ (Haislip et al. 2005b). This photometric redshift was confirmed and refined ($z = 6.10^{+0.37}_{-0.12}$) by Antonelli et al. (2005) using the 8.2-m Very Large Telescope observations in $I$, $J$, $H$, $K$ bands. A spectroscopic redshift, obtained with the 8.2-m Subaru Telescope, was later reported by Kawai et al. (2005), $z = 6.295 \pm 0.002$. For comparison, the most distant burst observed in the pre-Swift era was GRB 000131, whose redshift, measured with ESO-VLT, was $z = 4.50$ (Andersen et al. 2000).

Being the most distant GRB with a measured redshift, we discuss here the properties of GRB050904 in some detail. The idea is to understand whether such distant events exhibit properties which are different from the bulk of the observed population at $z \sim 1$–2. The X-ray spectra and light-curve of this burst are given by Watson et al. (2006b). The XRT light-curve extending up to 70 ks is presented in Figure 4.20. The X-ray afterglow fades by a factor larger than 1000 over the first day. Strong flaring activity is quite visible on the figure and its blow-up, this flaring is similar to that observed in other GRBs at early times (Burrows et al. 2005a). However, as can be seen in Figure 4.20 the light-curve does not settle into a power-law decay, continuing to be dominated by large variability. Concerning the spectrum, it softens from a power law with photon index 1.2 to 1.9, with possible evidence for large intrinsic absorption (Watson et al. 2006b).

The optical afterglow is discussed by Tagliaferri et al. (2005a). Its light-curve is not peculiar with respect to other GRBs. A break in the J-band light-curve is reported at $t_{bp} = 2.6 \pm 1$ day (observer frame). If this is a jet break Tagliaferri et al. (2005a) derive a beaming corrected energy $E_{\gamma} \sim (4-12) \times 10^{51}$ erg. This limit indicates that this burst is consistent with both the Amati and Ghirlanda relations (see Section 3.10). Another very interesting observation was reported by Boër et al. (2006). It concerns the detection of a very bright optical flare with the modest 25-cm TAROT telescope (Telescope Action Rapide pour les Objets Transitoires, Boër et al. 1999) during the prompt high-energy emission and the early afterglow. It is amazing and even unbelievable to imagine that optical emission from an object at $z \sim 6$ has been detected by a 25-cm telescope. Before this observation, such a statement would have been considered crazy by many astronomers, as it was in France at the end of the 1990s, when M. Boër and the authors of this book proposed TAROT to quickly observe the optical counterparts of GRBs. Five seconds after the alert distributed via the GRB Coordinate Network (GCN) TAROT started to observe the field and detected the counterpart. For very distant GRBs, the absorption of UV photons (with wavelengths shorter than 1216 Å) by neutral hydrogen in the host galaxy is
Figure 4.20. The early X-ray afterglow of GRB 050904 at $z = 6.29$ measured by Swift XRT (0.2–10 keV or 1.5–72.9 keV in the rest-frame). For comparison the flux of one of the most luminous X-ray sources known, the AGN SDSS J 1030 + 0534 is also shown; its flux has had to be multiplied by 100 to appear in the figure. Inset: the linear blow-up of the data from 10 to 70 ks illustrating the variability of the source at late times. It has been suggested that the XRT light-curve is dominated by emission from the central engine during the first 12 h of observations (Watson et al. 2006b).

detectable as a suppression of the flux at wavelengths shorter than $1216 \times (1 + z)$ Å; this is called the Lyman alpha cut-off. For GRB 050904, at $z = 6.3$, the Lyman alpha cut-off almost suppressed the afterglow below $1216 \times 7.3 \sim 8900$ Å. By chance, the first observations of a GRB with TAROT are always unfiltered, covering wavelengths from 3000 to 10 000 Å. The Burst Observer and Optical Transient Exploring System (BOOTES 1B), a 30-cm robotic telescope located in southern Spain, observed simultaneously with TAROT in the R band (from 5100 to 7500 Å), but no detection was achieved (Jelinek et al. 2005); the use of an R-band filter made the detection of such a highly redshifted object almost impossible due to the Ly$\alpha$ cutoff. This observation with TAROT was only the fourth detection of the early optical emission while the high-energy emission was still active. The others were GRB 990123 at $z = 1.60$ (Akerlof et al. 1999; Section 3.9), GRB 041219a (Vestrand et al. 2005) at an unknown
Figure 4.21. Optical (J band, lower curve) and X-ray (0.5–10 keV, upper curve) light-curves of GRB 050904 showing the end of the prompt emission and the transition to the afterglow (Boër et al. 2006). The optical emission was measured with TAROT from 80 s to 28 min after trigger. The X-ray light-curve is from Swift-XRT. The dotted line shows the backward extrapolation of the late afterglow measured 3 h after the trigger (Haislip et al. 2006).

redshift, and GRB 050401 at z = 2.9 (Rykoff et al. 2005). Boër et al. (2006) noted the similarity between GRB 990123 and GRB 050904: both optical flares are very bright, they have high peak energies in their restframe and large isotropic equivalent energy releases (∼few × 10^{54} erg). These properties and the similarity in the optical behavior of both events, at very different distances, are quite interesting. As we discuss in Chapter 6, the prompt optical emission is generally explained by reverse shock emission. Wei, Yan, and Fan (2006) suggested that this could also be the case for GRB 050904.

In the meantime, Swift slewed to the source direction, and the XRT started observing the field ∼160 s after the burst trigger. Figure 4.21 shows the comparison of the optical and X-ray (0.5–10 keV) light-curves. If the optical emission is attributed to the reverse shock, the temporal coincidence of the optical flash with an X-ray peak which is part of the prompt emission would be fortuitous, the X-ray flare being
attributed to the activity of the central engine (Wei, Yan, & Fan 2006). So, Wei, Yan, & Fan (2006) have proposed an alternative model where the optical flare would be produced by the synchrotron radiation of the electrons accelerated by the late internal shock due to a reactivation of the internal engine as was suggested for X-ray flares (see Chapter 7). This burst shows many flares (Figure 4.20) which can also be interpreted as late internal shocks related to central engine activity (Cusumano et al. 2006a). Wei, Yan, and Fan (2006) think that the late internal shock model is more favored because in this model the optical flash and the X-ray flare have the same origin, and this would explain their temporal coincidence.

Boër et al. (2006) have noticed the similarity of this burst with GRB 990123. Cusumano et al. (2006a) more generally indicated that the overall phenomenology of this very distant burst (a huge explosion in the very early Universe) is not peculiar with respect to other GRBs. This suggests that the mechanisms of GRB explosions are similar in the early Universe and today. Nevertheless, the possibility of different progenitors for GRBs at very high redshift is considered by Watson et al. (2006b); the question is natural because the gas which forms the progenitor star is certainly different. These authors consider that the high-energy output and the large-amplitude long-duration flaring are notable differences between this burst and typical GRBs. This might hint at an unusual progenitor in a different (i.e. low metallicity) environment.

As of early 2008, GRB 050904 was still the most distant GRB observed. The explosion that produced this burst occurred 12.8 billion years ago, about 900 million years after the Big Bang, when the Universe was very young. This period in the life of the Universe is extremely interesting because it is close to the transition between the dark ages and the era of re-ionization, which took place around \( z \sim 6–10 \) (see Becker et al. 2001, Spergel 2006, and Miralda-Escude 2003 for a review). At these early times the Universe was only about 6% of its current age. This is quite exciting because it means that stars were already forming, and that enough time was available for some of them to evolve and collapse into black holes. The question of whether the progenitors of such high redshift bursts can be due to Population III stars is open,\(^3\) with the difficulty for population III stars to lose their envelopes due to the low metallicity (wind-driven mass loss decreasing when the metallicity decreases). It seems, however, that even at this high redshift, galaxies were already enriched in metals. Besides, the spectroscopic redshift determined by Kawai et al. (2006) using a series of metal absorption lines indicates that the GRB is found in a region already enriched with metals. This suggests that metal absorption lines might be detected even from metal-free first-generation stars because their environment would be quickly self-polluted by pre-burst winds. Or, as suggested by Tagliaferri et al. (2005a), the progenitor might be a massive, non pristine star implying that star formation was already active at \( z > 6 \) and that metal enrichment had already started. This might be consistent with the fact that star formation rate seems to vary only slowly with redshift over the range \( 2 < z < 6 \). Le Fèvre et al. (2005) found a star formation rate

\(^3\) Population III is the name of the first generation of stars, which were created from pristine gas containing no metals.
\((10-100) \ M_\odot \, \text{yr}^{-1}\) for a sample of 970 galaxies with \(1.4 < z < 5\), indicating a star-formation activity higher than previously assumed at that early time.

More recently, Daigne, Rossi, and Mochkovitch (2006) using Monte Carlo simulations of the populations of long GRBs predicted their redshift distribution. The comparison with Swift data suggests the following conclusions: some properties of GRBs or/and GRB progenitors evolve with redshift. The main evolutionary effect could be the redshift dependence of the efficiency of GRB production by stars with a GRB comoving rate which increases above \(z \sim 2\) even if the star-formation rate flattens or decreases. Daigne, Rossi, and Mochkovitch (2006) suggest that this efficiency should be about 6–7 times larger at \(z \sim 7\) than at \(z = 2\), although it remains to understand why. So if this redshift dependence of the efficiency is confirmed it would mean that GRBs would not directly trace the star-formation rate. These questions of star-formation rates and metallicity in the early Universe are very crucial. They have been considered by Berger et al. (2007a) who studied the host galaxy of GRB 050904 with deep Hubble Space Telescope and Spitzer Space Telescope observations. They found that the host galaxy has a low mass (a few \(10^9 \ M_\odot\)) and is a modestly dusty starburst galaxy with a substantial star formation rate of \(~15 \ M_\odot \, \text{yr}^{-1}\) (see their Figure 4). Kawai et al. (2006) inferred a metallicity of about 0.05 \(Z_\odot\) from the \([S/H]\) value \([S/H] = -1.0\) measured in the afterglow absorption spectrum; this represents the metallicity in the star-forming region local to the GRB and not the average metallicity of all the HII regions of the galaxy. The evolution of the mass–metallicity \((M-Z)\) and luminosity–metallicity \((L-Z)\) relations (in the sense that galaxies of a given mass/luminosity have lower metallicities at progressively higher redshifts) has been observed from \(z = 0\) to \(z = 2\) (Savaglio et al. 2005, Erb et al. 2006), and it seems to continue at much higher redshift, i.e. \(z > 6\). The data for GRB 050904 (\(Z \sim 0.05 \ Z_\odot\)) are in good agreement with the empirical time evolution model of the \(M-Z\) relation derived by Savaglio et al. (2005) which indicates that for \(\log M \sim 9.5\) the expected metallicity at \(z = 6.3\) is about 0.1 \(Z_\odot\) (see their Figure 17). Berger et al. (2007a) emphasized the interest of GRBs to determine the metallicity at high redshift \((z > 5)\) in a domain where other techniques have difficulties due the weakness of the relevant emission lines. More generally the extension of the relations \(M-Z\) and \(L-Z\) beyond \(z \sim 5\) using GRBs and their afterglows (through GRB absorption spectroscopy) would be quite important to study the initial stages of mass-build-up and metal-enrichment.

The interest of such observations of high-redshift GRBs \((z > 5)\) has also been analyzed by Bromm and Loeb (2006), focusing on possible Population III progenitors for these high redshift GRBs. The interest for this GRB population is that it may represent a significant fraction of the GRBs \((\sim 10\%)\) detected by Swift. Jakobsson et al. (2006a) predicted that at least 7\% of GRBs originate at \(z > 5\) (more recently Jakobsson et al. 2006b estimated at least 5\%) and Daigne, Rossi, and Mochkovitch (2006) estimated 2–6 bright Swift bursts per year at \(z > 6\) based on a yearly rate of \(~100\) Swift GRBs. The predictions of Natarajan et al. (2005) and Gorosabel et al. (2004) are respectively 10\% and 2\%. Moreover at this early epoch Pop III stars would be viable progenitors for these long-duration GRBs as long as they can lose their outer envelopes through mass transfer to a companion star in a close binary system (but see Abel, Bryan, & Norman (2002), whose numerical
simulations suggest that all metal-free stars could be massive and form in isolation. Massive primordial stars would offer a natural explanation for the absence of purely metal-free low-mass stars in our Galaxy (Abel, Bryan, & Norman 2002). Jakobsson et al. (2006b) also indicate that it seems unlikely that Pop III stars could end their lives as GRBs. Nevertheless, if Pop III binaries were common and if a significant fraction of the massive stars are able to produce GRBs, which is today unknown, Swift might be the first observatory to probe Pop III star formation at redshift $z > 7$ (Bromm & Loeb 2006). In this case high redshift bursts would open a remarkably bright window for the study of otherwise dark (faint) ages (Abel, Bryan, & Norman 2002). Such studies are important because they allow us to access the earliest epoch of star formation where Pop III stars might represent the predominant population. These first stars in the Universe formed out of metal-free gas at redshift $z > 10$ would be predominantly very massive, above $100 M_\odot$, even if the final mass of the first stars remains unclear (Bromm, Coppi, & Larson 2002, Abel, Bryan, & Norman 2002), and with predicted high surface temperature. So they would be efficient sources of ionizing photons which can contribute to the re-ionization of the intergalactic medium. Moreover their short lifetime $\sim 10^6$ yr would lead massive Pop III stars to contribute largely and promptly to the initial enrichment of the intergalactic medium. High-redshift GRBs might be able to give observational constraints on the star-formation rate at these early epochs. Moreover, as a metallicity dependence of the GRB rate is expected in the collapsar model, future observations of GRB host galaxies might be used to test the relation between the average metallicity of the Universe and the GRB rate (Gorosabel et al. 2004, Natarajan et al. 2005). Swift and INTEGRAL-IBIS are particularly well suited to contribute in this cosmology field.

In Chapter 9 we will come back to the importance of high-redshift GRBs to probe the Universe when the first stars were formed.

Let us now consider two other GRBs whose peculiar behavior merits consideration here, even if they have not the same impact as the two GRBs we have just analyzed.

### 4.10.3 GRB 061007: the brightest GRB to date

This is an interesting case because it is one of the brightest GRBs to date\(^4\) with an exceptionally luminous afterglow with a V-band magnitude $< 11$ at 80 s after the prompt emission (Schady et al. 2007). This early optical flux is very close to the champion GRB 990123 (Akerlof et al. 1999), but for this burst the afterglow decay is different (Figure 4.22; Schady et al. 2007). The light-curves (V band, X-ray, and γ-ray) are remarkably parallel and without any break up to 150 ks for the optical light-curve and up to $10^6$ s at least for the X-ray light-curve. The decay indices for these light-curves are $\alpha_{UVOT} = 1.64 \pm 0.1$, $\alpha_{XRT} = 1.66 \pm 0.1$, and $\alpha_{BAT} = 1.54 \pm 0.12$.

This behavior, which started very early, $\sim 100$ s after the GRB trigger, is very different from other GRBs. The lack of any break requires an extremely large kinetic

\(^4\) Here again, this burst has been outshone by GRB 080319B whose prompt optical emission has reached magnitude 5.3 for a few seconds (see, e.g., Racusin et al. 2008).
Figure 4.22. Light-curves of the afterglow of GRB 061007 in all six UVOT filters, in X-rays (0.3–10 keV) and in gamma-rays (15 keV–350 keV). The left-hand axis corresponds to the V band light-curve and the count rate axis on the right applies to the X-ray and γ-ray light-curves (from Schady et al. 2007). The top panel shows a single UV/optical light-curve created from all the UVOT filters.
energy or a highly collimated outflow. Schady et al. (2007) suggested an early-time jet break occurring within 80 s of the prompt emission. This time leads to an upper limit on the jet opening angle $\theta_j \sim 0.8^\circ$, much narrower than the value $\theta \sim 4.7^\circ$ derived from the optical light curves by Mundell et al. (2007), which they consider a minimum value. Of course with such a narrow jet, less than 1°, there is no problem for the energy budget, reducing the isotropic equivalent energy of the burst to $E_e \sim 10^{50}$ erg. This unusual and remarkably simple decay of the light-curves was also analyzed by Mundell et al. (2007). They used optical observations obtained with the 2 m robotic Faulkes Telescope South (FTS) covering the period from 137 s to 3 days. They suggested that the afterglow emission started as early as 30–100 s and was contemporaneous with the on-going variable prompt emission from the central engine originating from a physically distinct region dominated by the forward shock. So these authors considered that the optical, X-ray, and late $\gamma$-ray emission of the expanding fireball is dominated by synchrotron emission from a forward shock whose cooling frequency lies above the X-ray band and whose typical synchrotron frequency $\nu_m$ is below the optical band from the very beginning.

The lack of a bright early-time optical flash and late-time radio flare is explained by Mundell et al. (2007) because the typical frequency of the reverse shock emission would already be in the radio domain at early times. A low $\nu_m$ would originate from small values of the microphysical parameters $\varepsilon_e$ and $\varepsilon_B$. The X-ray afterglow of this GRB does not have the complicated behavior reported for a large fraction of GRBs seen by Swift (see Section 4.8.1), so its simple behavior, its well-characterized light-curves and well-determined peak energy are very favorable to test the different spectral-energy correlations: Amati, Ghirlanda, Liang–Zhang. This has been done by Mundell et al. (2007) who concluded that this event is fully consistent with the $E_{\text{peak}}–E_{\text{iso}}$ correlation (Amati) but is an outlier to the $E_{\text{peak}}–E_e$ (Ghirlanda) and $E_{\text{peak}}–E_{\text{iso}}–t_h$ (Liang–Zhang) correlations (deviating by more than 3$\sigma$) with its X-ray afterglow continuing without a jet break until $t > 10^6$ s (but see Schady et al. (2007) who consider a very early-time jet break not observed but possibly occurring within the 80 s of the prompt emission before the first XRT observation). The existence of a population of objects similar to this burst, which are inconsistent with these correlations is evoked by Sato et al. (2007) and Willingale et al. (2007) (see Section 9.1 for a more detailed discussion of these correlations).

### 4.10.4 GRB 060614 a nearby long-duration GRB without supernova signature

GRB 060614 is interesting because it is a peculiar, nearby long-duration GRB not associated with a supernova. Let us review the main characteristics of this burst.

This GRB was bright, with duration $T_{90} \sim 102$ s observed by BAT (Figure 4.23; Gehrels et al. 2006). The inset of Figure 4.23 gives an expanded view of the first episode, which lasted $\sim 5$ s. The spectrum of this part of the burst, a power-law of index $\alpha_1 = 1.63 \pm 0.07$, is significantly harder than the spectrum of the following 120 s of the burst, a power-law spectrum of index $\alpha_2 = 2.02 \pm 0.04$. Similar two-component emission structures (a short hard spike followed by softer and longer emission) were reported in several Swift and HETE-2 short-duration GRBs. Norris
and Bonnell (2006) searched in the BATSE database and found three GRBs with a bright soft component following a short spike, as in GRB 060614, representing about 1% of the short-duration BATSE GRBs. For Swift and HETE-2, 4 among 16 short-duration GRBs present this weaker soft component. For Konus these events represent 10–25% of the short-duration GRBs (Mazets et al. 2004, Frederiks et al. 2004; see also the discussion in Section 4.5). As suggested by Gehrels et al. (2006) this behavior of some short GRBs might be the rule. These authors carefully analyzed the temporal behavior of this burst and particularly the temporal lags between the light-curves in different energy bands. Their conclusion is that it is difficult to determine unambiguously which category GRB 060614 falls into.

After the GRB detection the spacecraft slewed to the burst position in 90 s. The X-ray emission 0.3–10 keV was initially among the brightest detected by Swift. A break in the optical light-curve is suggested at $1.39 \pm 0.04$ d; if it is a jet break it
implies a narrow jet $\theta_j \sim 12^\circ$ (Della Valle et al. 2006) and suggests an energy of $10^{59}$ erg in the jet, comparable with that reported for Swift short GRBs. This GRB was localized to the outskirts of a relatively nearby faint dwarf galaxy at a redshift $z \sim 0.125$ (Price, Berger, & Fox 2006). A location at higher redshift ($1.44 < z < 1.71$) which would explain why no SN was visible was suggested by Schaefer and Xiao (2006). These authors think that the low redshift for this burst is due to the selection of the brightest galaxy near the line of sight, and they consider that this galaxy is just a random foreground source unrelated to the burst. But this suggestion of a chance alignment has been excluded (Gehrels et al. 2006, Gal-Yam et al. 2006). This discussion shows the difficulty of classifying such bursts. Nevertheless, the total duration $>100$ s puts this burst in the category of long GRBs, so it is natural to search for a supernova in its light-curve because previous long GRBs at such low redshifts showed clear signatures of underlying supernova explosions (except some XRFs, see Section 4.4.2). Surprisingly no supernova was detected and very tight upper limits were reported by three groups (Fynbo et al. 2006, Gal-Yam et al. 2006, and Della Valle et al. 2006). For instance Della Valle et al. (2006) observed this burst with the European Southern Observatory 8.2-m Very Large Telescope (ESO-VLT). The host galaxy revealed on-going star formation and the presence of young massive stars but this galaxy is fainter than most GRB host galaxies. The observations obtained in the R band up to 65 d after the burst did not reveal the emergence of a SN component (Della Valle et al. 2006, see their Figure 2). These authors estimated that the source was affected by less than 0.2 mag of extinction in the observed R band. They concluded that any SN associated with GRB 060614 had to be at least 100 times fainter at optical wavelengths than the supernovae associated with ten GRBs summarized in Ferrero et al. (2006). Gal-Yam et al. (2006) arrived at the same conclusion, noting that any putative SN component would have to be more than 100 times fainter than the faintest event previously known to be associated with a long GRB (for instance like GRB060218–SN 2006aj, considered in Section 4.9.1) and in fact fainter than any SN ever observed with an amount of synthesized $^{56}\text{Ni} \sim 5 \times 10^{-4} M_{\odot}$. These authors concluded that their HST data indicate that this GRB was not associated with an event powered by radioactivity, like all known supernovae. Fynbo et al. (2006) considering this burst and GRB 060505 at $z = 0.089$ (a faint burst with a duration of 4 s), both located in star-forming galaxies concluded that neither GRB 060505 nor GRB 060614 were associated with significant SN emission down to very faint limits, hundred of times less luminous than the archetypal SN 1998bw. They also indicated that even the faintest Type Ic SN (SN 1997ef) would have been easily detected in their observations (Figure 4.24; Fynbo et al. 2006). These results were obtained after the level of galactic extinction had been checked and found very low in both cases. Moreover the authors looked for dust obscuration at the location of the source, and they ruled out significant obscuration of the source in the host galaxy for the two bursts. So their conclusion, similar to the previous ones is that any SN associated with these two GRBs must have been not only fainter than any SN previously associated with a GRB or XRF but also substantially fainter than any non-GRB-related Type Ic SN seen until now. A lot of excitement has followed this puzzling result.
Figure 4.24. Light-curves of supernovae SN 1998bw, SN 2002ap, and SN 2006aj as they would have appeared at the redshift of GRB 060505 (a) and at the redshift of GRB 060614 (b). It clearly appears that neither GRB 060505 nor GRB 060614 were associated with significant SN emission down to very faint limits hundreds of times less luminous than SN 1998bw. These two GRBs illustrate the case of no-supernova, long GRBs, quite different from previously observed nearby long duration GRBs (Fynbo et al. 2006).

On the one hand the duration of GRB 060614 is consistent with that of a long burst, even though it has been argued that perhaps it is not a really typical long-duration GRB. On the other hand the lack of an associated SN is inconsistent with the progenitor expected for long GRBs: a massive, rapidly rotating star which undergoes a core-collapse SN explosion (collapsar).
The classical scenario invoked for long and short GRBs with distinct physical origins (long GRBs from SN/collapsar and short GRB from binary mergers—see Chapter 8) seems to be in difficulty and perhaps has to be revisited. Many suggestions have been proposed to explain these results. Gehrels et al. (2006), noticed that short/hard bursts with extended activity resembling long soft bursts have been observed (see also Section 4.11) and such bursts are predicted by some models of massive stars when the pre-explosive mass loss is unusually high and the jet energy is low (Zhang, Kobayashi, and Mészáros 2003). To explain the very low upper limits on $^{56}$Ni production, it is possible that $^{56}$Ni which has been produced initially falls back onto the collapsed core in a weak SN explosion. Fryer, Young, and Hungerford (2006) studied the explosive nucleo-synthesis from GRB progenitors in the two cases of BH formation, either directly with no SN explosion or through the fall-back of considerable material onto a proto-neutron star formed in a weak SN explosion followed by the ejection of a part of this material in a second explosion (Fryer 1999). This second scenario would be the most common path to BH formation (Fryer & Kalogera 2001). It is interesting to note that the $^{56}$Ni yield by a BH formed after a considerable delay (longer than 1 s) is much lower than in the case of direct BH collapses or BH formed shortly after a weak explosion. The trend is for a decreasing Ni production with an increase in the delay between the weak and the strong explosion (Fryer, Young, & Hungerford 2006). Gal-Yam et al. (2006) also considered the possibility of weak SN explosions: besides the classical collapsar there would exist a category of progenitors synthesizing very little $^{56}$Ni, which do not produce optically luminous SNe. In this case the massive progenitor star is probably different from the progenitors of typical long GRBs.

Another way to explain the lack of a supernova would be to associate such GRBs with low-mass, long-lived progenitors which might be similar to those producing short GRBs, but with no supernova; Fynbo et al. (2006) call them SN-less GRB progenitors suggesting a new phenomenological type of massive star death. Della Valle et al. (2006) evoke a similar possibility, based on a class of very faint core-collapse SN discovered in the local Universe (Pastorello et al. 2004). They are SN of Type II with absolute magnitude in the range $-13 > M_\odot > -15$, significantly lower than typical core-collapse supernovae. These SN eject a very small mass of $^{56}$Ni during the explosion: a few $10^{-3} M_\odot$, at least a factor of 10 lower than the average for core-collapse supernovae and show very small expansion velocities (Della Valle et al. 2006). These properties seem consistent with data obtained on GRB 060614. These SN would correspond to the collapse of massive stars but with a very small explosive energy, which explains why most of the $^{56}$Ni produced falls back onto the compact stellar remnant. But even if such SNe are able to produce a BH, which seems the necessary condition to produce a GRB, an interesting question is how the stellar envelope is liberated to allow the relativistic jet source of the GRB to emerge out of the star. Clearly this type of progenitor remains to be explored.

Another possibility would be a SN explosion appearing significantly before the GRB, as proposed in the supranova model of Vietri and Stella (1998) (see Section
8.4). Such a model would avoid the SN explosion at the time of the GRB, justifying the impossibility of detecting it.

Yet another possibility, which avoids a supernova explosion completely, is to postulate that the emission of GRB 060614 occurs in a NS–NS merger (Gehrels et al. 2006). In this case the accretion has to be longer than in a typical NS–NS merger for which the lifetime of the accretion disk around the newly born BH is expected to be quite short, \( \sim 0.1 \text{ s} \) (Narayan, Piran, & Kumar 2001). However, as shown by Faber et al. (2006), during the BH–NS merger a significant amount of mass can be ejected to a bound orbit and fall back into the BH on timescales which might be larger than 1 s. Another way would be to consider a WD–NS merger. King, Olsson, and Davies (2007) showed that such mergers are likely to produce long-duration GRBs, in some cases definitely without an accompanying SN. King (2006) gives the conditions on the mass of the progenitors to have dynamically unstable mass transfer. This class would have a strong correlation with star formation and occur close to the host galaxy. These GRB progenitors would be situated in any type of host galaxy (King, Olsson, & Davies 2007). Such mergers might make a significant contribution to the GRB rate. This conclusion does not disagree with the fact that the fraction of SN-less GRBs could be substantial because, of the six long-duration GRBs or X-ray flashes known at low redshift \( (z < 0.2) \); two have no associated SN: GRB 060505 and GRB 060614 (Fynbo et al. 2006). To conclude, if these bursts can be explained by the coalescence of two compact objects, we are just in the situation of short bursts with a soft energy tail lasting tens of seconds, and it is natural to have no associated supernova.

So, various progenitors of long GRBs without SN are possible but we cannot completely exclude the possibility of a SN explosion that would be very faint and not detectable even at low redshift, or the possibility that the source is really at higher redshift (but this is less probable). In any case, the commonly accepted scenario with two kinds of progenitors for short and long GRBs, largely discussed in Chapter 8, might be more complicated. Perhaps a new GRB classification scheme is needed, as suggested by Gehrels et al. (2006), but many more events with this property (long GRBs with no SN) have to be observed to justify the introduction of a new classification. First, it would be important to clarify the situation of short GRBs followed by a softer energy tail or to avoid the classification of GRBs solely on their duration. Zhang et al. (2007) proposed to abandon the terms ‘short’ and ‘long’, and by analogy with SN classification, to use the alternative terms of Type I and Type II GRBs (see also Bloom & Prochaska 2006). Type I GRBs would be associated with old stellar population, like Type Ia SNe appearing in galaxies of all sorts, with their likeliest progenitor candidates being compact star mergers. Type II GRBs would be associated with young stellar population, appearing in star-forming regions of their host galaxies, and likely produced by core collapse of massive stars, like Type II and Type Ibc SNe. If GRB 060614 is a Type I GRB its progenitor is a compact star merger, and the absence of SN is normal. But there is still a problem: how to make extended high-energy emission from a merger type GRB. This is also a question raised by the X-ray flares seen in the afterglows of short GRBs (see Chapter 7).
Clearly, the discovery of GRB 060614 without SN signature and its consequences are among the first-class results of Swift, but before abandoning the commonly accepted classification more bursts with such properties (low redshift and no SN signature) are needed. Let us now go to another major result of Swift: the detection of the afterglows of short GRBs, which was eagerly expected after the absence of BeppoSAX results on short GRBs.

4.11 AFTERGLOWS OF SHORT GRBS

In the first year of Swift operation 11 short hard GRBs (SHB) were accurately localized within an hour; this was much more than any previous mission. Swift provided the discovery of the first X-ray afterglow of an SHB (GRB 050509B), and the probable identification of its host galaxy and redshift. In July 2005 the breakthrough occurred with the detection of the optical afterglows of GRB 050709, localized with HETE-2, and GRB 050724, localized with Swift.

A summary of the first localizations of SHBs with Swift has been given by Roming et al. (2006b):

(a) Four SHBs localized by BAT have no corresponding X-ray, optical or radio counterparts: GRB 050202 (Tueller et al. 2005), GRB 050906 (Krimm et al. 2005), GRB 050925 (Holland et al. 2005), and GRB 051105A (Mineo et al. 2005).
(b) Three SHBs have X-ray afterglows but no optical or radio counterparts: GRB 050509B (Gehrels et al. 2005), GRB 050813 (Retter et al. 2005), and GRB 051210 (La Parola et al. 2006).
(c) Four SHBs have X-ray and optical afterglows: GRB 050724 (Barthelmy et al. 2005b), GRB 051221A (Burrows et al. 2006, Fan, Piran, & Xu 2006, Soderberg et al. 2006a, Jin et al. 2007), GRB 051227 (Barbier et al. 2005), and GRB 060313 (Roming et al. 2006b).

The following discussions are limited to the conclusions obtained through the analysis of SHBs observed in the year 2005. At the time of this writing, interesting new results have been obtained on bursts observed in 2006; the reader can consult Berger et al. (2007b) and Nakar (2007) for an overview of these results. Coming back to SHBs observed in 2005, only for a limited number of them were the determination of the redshift and the identification of the host galaxy possible:

- GRB 050509B was the first Swift SHB which could be (most probably) associated with a host galaxy: an elliptical at \( z = 0.225 \).
- GRB 050709 (HETE-2) presented us with the first detection of an optical afterglow. The host galaxy at \( z = 0.16 \) is a late-type star-forming blue dwarf galaxy. The association of the SHB with a supernova is rejected.
- GRB 050724 is associated with an elliptical galaxy at \( z = 0.258 \); this is the first SHB with a radio afterglow.
• GRB 051221, at $z = 0.54$, took place in an actively star-forming galaxy which contains a significant population of old stars. No bright supernova is observed.

These SHBs were, at the beginning of 2006, the most interesting because their host galaxies and redshifts had been identified. This relatively small sample of SHBs allows us to draw the following conclusions:

• As predicted by models invoking mergers of compact objects, the afterglows of SHBs appear to be fainter than the afterglows of long-duration bursts; it also seems that short bursts are only weakly collimated.

• Short GRBs are usually not associated with star-forming regions, even though they appear in a large variety of galaxies (Prochaska et al. 2006); elliptical galaxies dominated by old stellar population, as well as star-forming galaxies with a star-formation rate which is lower than the average star-formation rate of long GRB hosts (Berger 2006). These first identifications showed the diversity of the host galaxies, allowing us to say that the stellar population in SHB hosts is generally older than in the hosts of long GRBs (Bloom & Prochaska 2006).

• The location of SHBs in their host galaxies also seems to differ from long GRBs (see for instance Figure 2 in Bloom & Prochaska 2006), but today it is difficult to say that they are substantially different. The prudent conclusion of these authors is that SHBs appear to be more diffusely positioned around galaxies than long GRBs.

• It also appears that most secure redshifts are rather low: $z < 0.5$ with a median value $z = 0.24$ (Bloom & Prochaska 2006), compared with the average $z = 2.8$ for long GRBs (Jakobsson et al. 2006b). So, Guetta and Piran (2006), using the new constraints from this small sample of five Swift/HETE SHBs with redshifts, concluded that the sample is incompatible with an SHB population that follows the star-formation rate, but compatible with a distribution of delay times after the SFR, favoring the merger model. However, the sample is too small to claim that this conclusion is robust. It depends on the global redshift distribution which is still an open issue.

The SHB sample was increased in 2006 thanks to optical observations of seven new SHBs, five of them discovered in 2006 plus GRB 051210 and GRB051227, leading to quite different conclusions (Berger et al. 2007b). These observations were obtained with Gemini, Magellan, and the Hubble Space Telescope. Based on these new observations, Berger et al. (2007b) found that the most probable host galaxies of many SHBs are faint and that the new hosts likely resided at higher redshifts, $z > 1$. They concluded that there is a ‘new population’ of SHBs, and suggested that 1/3 to 2/3 of all SHBs originate at higher redshifts than reported after the observations of 2005. Of course, this new trend has to be confirmed by more SHB detections and identifications.

Based on the more robust identifications and redshift determinations made in 2005, it seems that these observations favor the binary merger model invoking the
coalescence of two compact objects, particularly BH–NS mergers. Analytical and numerical arguments suggest that the disruption and swallowing of the NS by the BH may lead to complex and more extended accretion than for double NS mergers. Indeed Davies, Levan, and King (2005) showed that the final merger of a BH–NS system is complex with different steps before the NS is shredded and forms a disk sufficiently massive (~ 0.1 $M_\odot$) for GRB production. The entire process has a relatively long duration, a few seconds or many thousands of orbital periods. But if this type of progenitor seems to distinguish the class of short bursts from the class of long-duration GRBs, which are best explained by the collapsar model, it should be noted that the SHB X-ray afterglows show similarities with typical X-ray light-curves of long-duration GRBs. The presence of X-ray flares appears to be a common property of both classes of bursts. Evidence for a jet break has also been reported for SHBs (Burrows et al. 2006, Panaitescu 2006b). So, the ejecta appears to be collimated but with opening angles that would be larger than the median value of ~5° observed for long GRBs. Such similarities can be understood if we recall that in both cases the pre-burst system is a BH surrounded by a torus of matter. Another intriguing point concerning SHBs is the fact that for some of them the hard $\gamma$-ray pulse is followed by a soft tail lasting several tens of seconds. These tails might be the beginning of the afterglow or they could be X-ray flares, like those following long GRBs.

In conclusion, Swift (and HETE-2 for a few bursts) has allowed major advances in the understanding of SHBs. The identification of a few host galaxies seems to confirm that SHBs are a distinct class, as was first established through their intrinsic spectral and temporal properties. These host galaxies seem to prove that the SHBs have progenitors which are distinct from those of long-duration GRBs. Old population objects compatible with binary mergers, NS–BH or NS–NS, are today valid candidates for SHBs, even if the sample of SHBs which has been analyzed is too small to be sure of this classification.

As we have seen, the redshift distribution of SHBs is not at all determined. Although the first identifications made in 2005 supported a small median redshift $z = 0.24$, the results of Berger et al. (2007b) for seven bursts support the existence of a population of SHBs with significant higher redshifts. The characterization of SHBs appears complex, as is the characterization of long GRBs with the discovery of two nearby long-duration GRBs without any ‘trace’ of SN. This puzzling discovery of Swift may cast doubts on a classification of GRBs into two classes but here also it is too early to be more specific. In this context, the classification proposed by Zhang et al. (2007) with Type I and II GRBs is attractive because it avoids this notion of short and long GRBs with a distribution which is now blurred.

### 4.12 CONCLUSION

In this chapter we have stressed some major results obtained with two missions, HETE-2 and Swift, entirely devoted to the study of the prompt and afterglow emission of GRBs. We have also reported some results obtained with INTEGRAL, which like Swift, is still operating in 2008.
HETE-2 provided fast positions, which have encouraged organizing follow-up campaigns with the most powerful telescopes. The optical follow-up of GRB 021004 from 193 s to 3 months after the burst is a good example of this capability of HETE-2 to disseminate the GRB position quickly (within tens of seconds). The importance of arriving early after the GRB trigger has been also illustrated by the observation of GRB 021211 which demonstrated that bursts which can be dark when observed hours after the trigger may be bright in the first minutes. HETE has also contributed to the very important search for SNe associated with GRBs a decisive step in determining the nature of the progenitors of long GRBs. GRB 030329, one of the closest GRBs, with a redshift $z = 0.1685$, was associated with a SN which was called SN 2003dh, confirming the association of long GRBs with the collapse of massive stars. Another breakthrough of HETE-2 was the discovery of the first optical afterglow of a short GRB. This short burst also revealed the presence of a long-lasting ($\sim 100$ s) soft bump following the initial spike. With Swift we have seen that the second half of year 2005 was very prolific with the discovery of several afterglows associated with short GRBs. Another field where HETE-2 was particularly successful was the discovery of many X-ray flashes. Thanks to its broad energy range HETE-2 was able to localize X-ray flashes as well as GRBs. The study of the sample of XRFs obtained with HETE-2, in addition to the first Swift observations of XRFs, confirmed that X-ray flashes are in fact soft GRBs. The possible nature of the XRFs has been discussed through the analysis of several models. But in the end it seems that the emission of relativistic jets by dying massive stars followed by the production of high-energy photons might be a general behavior. Hence, depending on the energy involved, a classical GRB, an XRF, or a low-luminosity GRB might be produced.

Before the analysis of the major breakthroughs of Swift we discussed some interesting contributions obtained with INTEGRAL and its quite fruitful collaboration with XMM-Newton. These two ESA missions in similar orbits allowed XMM to follow many INTEGRAL GRBs at X-ray wavelengths.

Concerning Swift, the unique ability of the satellite to slew rapidly and autonomously and to track GRBs and their afterglows with a minimum delay (tens of seconds) with powerful instruments allowed major breakthroughs in the early X-ray afterglows which were out of reach for BeppoSAX. This led to the discovery of new phases of the early X-ray afterglow: a steep decay following the prompt phase and, particularly interesting and quite new, a shallow phase not anticipated in the standard model based on internal and external shocks. Moreover, X-ray flares which can be very intense were discovered in a large fraction of X-ray light-curves, about half of them. In addition to these fundamental discoveries on the early X-ray afterglows we have reported the discovery of some quite exciting GRBs such as:

- **GRB 060218**, the first low-redshift X-ray flash.
- **GRB 050904**, a distant GRB with a redshift of $z \sim 6.3$. The explosion took place about 900 million years after the Big Bang, when the Universe was still very young.
- **GRB 061007**, a very bright GRB, now outshone by GRB 080319B.
- **GRB 060614**, a nearby ($z \sim 0.125$) long-duration GRB without supernova light.
This burst and GRB 060505, which is similar, have given rise to great debates on the different ways to explain a lack of supernova light which is quite unusual for nearby long GRBs. But, in the end, many reasons have been found to explain the absence of SNe, without implying any need for a new classification of GRBs.

To close this long list of first-class results, we have presented the first observations of the afterglows and hosts of several short GRBs. Even though the sample of short GRBs with afterglows is still small, some interesting conclusions have been obtained. An important one is the confirmation that short and long GRBs (or Type I and II to avoid a discrimination by their duration) are really classes of GRBs with different progenitors. The properties of short GRBs (or Type I) and their hosts support binary mergers (NS–NS or BH–NS) as the most probable progenitors.

4.13 REFERENCES


Holland, S., et al. (2002a) GRB 021004: optical photometry at 1.5 days, *GRB Coordinates Network* 1585, 1.


Spergel, D. N. (2006) Cosmic microwave background: probing the Universe from $z = 6$–1100: the Universe at $z > 6$, 26th Meeting of the IAU, Joint Discussion 7, 17–18 August 2006, Prague, Czech Republic, JD07.


5

Gamma-ray bursts—fireballs and blastwaves

The observational discoveries described in the previous chapters have led over the years to the development of the GRB ‘standard model’ that describes the main properties of gamma-ray bursts with standard physics applied to somewhat ‘exotic’ objects. In this chapter we outline the most popular model to explain GRBs and their afterglows. This model, with its basic ingredients, has very often been called the standard model.

In Section 5.1, we discuss the behavior of a relativistic fireball during the free expansion phase, particularly the evolution of its bulk Lorentz factor. We explain why a simple fireball cannot account for the non-thermal emission observed in GRBs, even if its expansion is ultra-relativistic, leading to the necessity to consider shocks associated with this fireball.

After this short introduction of the ‘standard model’, invoking internal and external shocks, we analyze some basic elements needed for the understanding of GRBs and their afterglows (Section 5.2): the shocks associated with ultra-relativistic fireballs, the content of the fireball (particles and magnetic fields and their characterization), the acceleration of electrons in relativistic shocks, and the resulting spectrum of energetic electrons. Finally, having energetic electrons and magnetic fields, we discuss how GRB photons are produced by synchrotron or and synchrotron + inverse Compton emissions. The photon spectrum of the prompt emission and of the afterglow are briefly described with some useful characteristics. The possible role of inverse Compton scattering is discussed, taking into account the fact that the standard model is based on photon emission from synchrotron radiation only.

After this description of the basic ingredients of the standard model, Section 5.3 concentrates on the internal shock scenario introduced to explain the prompt GRB emission.
5.1 FIREBALLS

5.1.1 Relativistic expansion and pair opacity

Since GRBs are produced at cosmological distances their fluxes imply huge energies, which can reach $10^{54}$ erg if the emission is isotropic. Astrophysicists usually assume that the fastest variations measured in an astronomical source constrain its size $R$, because all the fluctuations shorter than the light-crossing time of the source will be smeared out by propagation delays within the source. Defining $\Delta t$ to be the most rapid variations of the luminosity of the source, this constraint leads to $R < c\Delta t$. This is generally not a very strong constraint, but it becomes very significant for cosmological gamma-ray bursts which show millisecond variability. In the absence of special relativistic effects (see below), this implies typical emitting regions of several thousand kilometres, where $10^{51} - 10^{52}$ erg (more than the energy emitted by the Sun during its entire life) must be released in a few seconds. These huge energies within a very small volume imply the formation of an $e^+, e^-, \gamma$ fireball.

This concept of a fireball was introduced for the first time by Cavallo and Rees (1978). Such giant explosions are also observed in supernovae with similar energies, but in the case of GRBs the huge energy has to be released into a smaller volume and in a shorter time; therefore the source must be extremely opaque to pair creation and it cannot emit non-thermal radiation. The question of pair creation has been central over the history of gamma-ray bursts because it provides an extremely efficient way to suppress photons above $\sim 1$ MeV. The observation of a significant flux above $\sim 1$ MeV in several GRBs (see, for example, Matz et al. 1985) showed that pair creation was not predominant in these sources. This implied a low density of photons at the sources, which could only be accommodated by nearby sources (at a maximum distance of few hundred parsecs), given the high fluxes measured at the Earth (Schmidt 1978).

Of course this solution does not work for cosmological GRBs, and we must have fireballs which are initially very opaque to pair creation. Due to this large opacity the fireballs are expected to expand relativistically (Paczynski 1986, Goodman 1986). The highly relativistic expansion of the fireball modifies its radiation: the observed photons are blueshifted (an X-ray photon in the fireball becomes a $\gamma$-ray photon in the observer frame) while the observed timescales are shortened. Relativistic expansion thus provides a very efficient way to reduce the rate of pair creation: in the moving source frame the photons are softer by a factor $\Gamma$ (the Lorentz factor of the relativistic flow), and they come from a region whose typical size is $\Gamma^2 c \Delta t$ (instead of $c \Delta t$ for a stationary source), so the density of photons is thus reduced considerably. As an example, for a typical GRB lasting 10 s, the typical size of the radiating region is about $3 \times 10^{15}$ cm if the Lorentz factor is 100 (see the discussion of the afterglow in Section 6.5). Finally, relativistic beaming implies that we observe only a small fraction $1/\Gamma$ of the source, whatever the opening angle of the emission. Therefore the relative angle at which the photons collide must be less than the inverse of the bulk Lorentz factor $\Gamma^{-1}$ of the relativistic flow; this reduces drastically the effective pair production if $\Gamma$ is large. These three effects lead to a large decrease of the optical
depth for pair creation. The first modifies it by a factor $\Gamma^{-2\alpha}$ where $\alpha$ is the photon’s index of the observed photon spectrum ($dN_{\gamma}/dE \propto E^{-\alpha}$), and the second modifies the photon density estimate by a factor $\Gamma^{-4}$ and influences the optical depth by a factor $\Gamma^{-2}$. Taking into account these effects Piran (1999) and Lithwick and Sari (2001) found that for $\alpha \sim 2$ the source becomes optically thin if $\Gamma$ reaches $\sim 100$. The calculations show that with such Lorentz factors the pair opacity is below unity before the fireball starts to radiate a significant fraction of its energy.

### 5.1.2 Basic fireballs

As we have seen, fireballs are essentially dynamic objects whose properties evolve quickly with time. This section discusses the evolution of standard fireballs (which is indeed very similar with the evolution of the first minutes of the Universe), and the modifications that have been brought in to explain GRBs.

A fireball can be characterized by its initial energy $E_0$. The energy-to-mass ratio in the initial fireball is $\eta = E_0/M_0c^2$ with $M_0 \ll E_0/c^2$, $M_0$ represents the baryon loading of the fireball and $\eta$ the mean energy per baryon. The initial radius of the fireball is $R_{in}$. Initially inside this radius $R_{in}$ the particles have a bulk Lorentz factor $\Gamma \sim 1$ and random isotropic velocities with a mean Lorentz factor $\gamma \sim \eta = E_0/M_0c^2$. The initial optical depth being extremely high, radial expansion is the consequence of the highly super-Eddington luminosity and the internal energy can only be converted into bulk kinetic energy. A phase of acceleration begins and the outward expansion of the particles leads their velocity vectors to be confined inside an increasingly smaller angle $(R/R_{in})^{-1}$ along the radial direction, the velocity distribution staying isotropic in the comoving frame. The rest-frame temperature $T'$ and the bulk Lorentz factor $\Gamma$ of the exploding fireball can be derived from thermodynamics and energy conservation.

Adiabatic expansion gives $T'V'^{\gamma-1} = cst$, where $V'$ is the volume of the emitting region and $\gamma = 4/3$ is the adiabatic index of the gas, $V' \propto R^3$ in the initial stage, so $T' \propto R^{-1}$. As the total energy is constant, $T' \propto \Gamma^{-1}$, and $\Gamma \propto R$, the bulk Lorentz factor of the expanding gas increases linearly with $R$ until it saturates at a value $R_{sat} \propto \Gamma_{max}R_{in}$, with $\Gamma_{max} \sim \eta \sim E_0/M_0c^2$, and $R_{sat} = \eta R_{in}$. During this phase the fireball thermal energy is adiabatically transformed into bulk motion energy. Beyond the distance $R_{sat}$ the shell coasts at constant $\Gamma = \Gamma_{max}$. Generally, the deceleration of the ejecta, due to its interaction with the surrounding medium, occurs after the bulk Lorentz factor $\Gamma$ has saturated, and the deceleration radius $R_d$ is larger than $R_{sat}$ (Mészáros & Rees 1993b, Mészáros, Rees, & Papathanassiou 1994).

As all the matter has moved with $v = c$ since the beginning, it is all piled up in a thin shell. In fact the thinness of the shell can be explained. The radially expanding shell has a width in the observer’s frame which is initially $\delta R \sim R_{in}$. The radial velocity spread causes a gradual spread of the radial width of the shell in the observer’s frame $\delta R/R \sim \delta v/v \sim \Gamma^{-2}$. The width of the shell in the observer’s frame is therefore $\delta R \sim R_{in}$ for $R < R_{\delta}$ and $\delta R \propto R\Gamma^{-2}$ for $R > R_{\delta}$, with $R_{\delta} \sim \delta R\eta^2$ being
the spreading radius (Mészáros, Laguna, & Rees 1993). When $\eta$ is large this radius is much larger than $R_{\text{sat}}$ because $R_{\text{sat}} \propto R_{\text{in}} \eta$. For instance for $R_{\text{in}} \sim 10^7$ cm and $\eta \sim 10^3$ this gives $R_\delta = 10^{13}$ cm.

The relativistically expanding fireball initially has $e^\pm$ pairs in equilibrium that dominate the scattering opacity, leading to very high optical depth. The pairs in the expansion fall out of equilibrium and recombine below a comoving temperature $T' \sim 17$ keV. Such fireballs present some difficulties in explaining gamma-ray bursts (Mészáros 2002):

- The radiation emitted at transparency is quasi-thermal, producing a thermal photon spectrum instead of the observed non-thermal power-law GRB spectra.
- The typical timescales over which the photons escape are comparable to that during which the flow becomes optically thin (milliseconds). This is too short with respect to most GRB durations.

To avoid these major problems, Rees and Mészáros were the first to introduce a new version of the fireball: the fireball shock model with external shocks (Rees & Mészáros 1992) and internal shocks (Rees & Mészáros 1994). Rees & Mészáros (1992) suggested that adding some baryon contamination in the fireball may solve the two problems mentioned above. Even a small amount of baryons ($10^{-7}$–$10^{-5} M_\odot$) entrained in the fireball would end up carrying the bulk of the fireball energy, transforming the initial radiative energy into kinetic energy. The interaction of the baryons with the surrounding medium, taking place in relativistic shocks, allows the energy of the fireball to be radiated away with a non-thermal spectrum (Rees & Mészáros 1992). For not too small baryon loads the density of pairs decreases to values which are much less than the density of ‘baryonic’ electrons associated with protons. For an initial black-body temperature of a few megaelectronvolts (at $R_{\text{in}}$), pair recombination starts at radii below the saturation radius. At this radius the optical depth has decreased but it is still larger than 1, and the radiation does not escape. A photospheric radius $R_{\text{ph}}$ is defined when the optical depth $\tau_{\text{ph}}$ reaches unity. Beyond $R_{\text{ph}}$ the radiation is able to escape but it can arrive at the observer only if the velocity vector of the radiating matter makes an angle smaller than $\theta \sim \Gamma^{-1}$ with the line of sight (Mészáros, 2006). These models are based on the assumption that ultra-relativistic outflows are likely to generate shocks and if these shocks occur after the fireball has become optically thin they would reconvert the bulk kinetic energy of the flow into internal energy of non-thermal particles and radiation. It is also important to have a sufficient number of particles, and especially electrons, which can be accelerated in the shocks and reach high energies. These particles will then radiate through synchrotron emission or combined synchrotron and inverse Compton emissions to produce the observed gamma-ray spectra. Various shocks are expected to take place in these models.

- **External shocks (ES):** the shock of the fireball on the external medium surrounding the source. This shock explains quite well the multi-wavelength afterglow radiation. In fact the external shock model anticipated the presence of afterglows

- **Internal shocks (IS):** These shocks are favored to explain the rapidly variable light-curves of the prompt gamma-ray emission. They occur inside the fast-moving ejecta when the time-varying outflow from the central engine leads to successive shells ejected with different Lorentz factors. Multiple shocks appear when faster shells overtake slower ones.

In the external shock the flow interacts with the external medium, whereas in the internal shock the flow is interacting with itself. This gives rise to an important difference since external shocks are highly relativistic, while the internal shocks are only mildly relativistic (Rees & Mészáros 1994). The condition for the shocks to occur when the fireball is optically thin is obviously more stringent for internal shocks, which take place closer to the central engine. They have to appear beyond the photospheric radius \( R_{\text{ph}} \), to allow most of the energy to come out as non-thermal radiation. Such a condition has consequences for the range of allowed values for \( \eta \), which must typically be between a few tens and a few hundreds (Mészáros 2006).

![Figure 5.1.](image-url)  
**Figure 5.1.** Schematic behavior of the jet Lorentz factor \( \Gamma \) in the different regions that can be identified in the standard model (adapted from Mészáros 2006). \( R_{\text{rad}} \) (\( r_s \) in the figure) is the distance defining the end of the acceleration phase, during which the bulk Lorentz factor \( \Gamma \) increases linearly with \( R \). \( R_{\text{ph}} \) characterizes the photospheric radius. \( R_{\text{in}} \) defines the end of the coasting phase with \( \Gamma \) staying constant. \( R_{\text{is}} \) and \( R_{\text{es}} \) are used to define the beginning of internal and external shock phases. The prompt \( \gamma \)-ray emission is produced by internal shocks at a typical distance \( R_{\text{is}} \). After \( R_{\text{es}} \), \( \Gamma \) decreases; \( R_{\text{es}} \) defines the deceleration radius and is often called \( R_{\text{dec}} \). It indicates the beginning of the afterglow. The dependence of \( \Gamma \) versus \( R \) on its decreasing phase is \( \Gamma \propto R^{-3/2} \) in the adiabatic case and \( \Gamma \propto R^{-3} \) in the radiative case, the medium being assumed to have uniform density. We have also indicated the typical values of these different radii, which can be found for instance in Mészáros (2006). This figure also indicates the typical radiations produced in the different regions. The production of thermal \( \gamma \)-rays at the photosphere is indicated by \( \gamma_{\text{th}} \), while \( \gamma \) indicates gamma-rays produced by internal shocks. \( \gamma, X, O, R \) characterize the afterglow emission by external shocks starting at \( R_{\text{es}} \).
Figure 5.1 (from Mészáros 2006) indicates the schematic behavior of $\Gamma$ versus $R$. This figure also shows where internal and external shocks appear and what kind of radiation emission can be associated with these regions. We also give typical order of magnitude values for these radii, the true expression being given in the literature (see, for example, Piran 2005, Mészáros 2006) and in the following sections.

This internal–external shock model does not depend on the nature of the central source. It is irrelevant whether the initial energy source is due to the coalescence of compact objects or to a collapsar produced by the explosion of a super-massive star (see Chapter 8 on GRB progenitors). The development of the internal–external shock model made it possible to explain many of the properties of GRBs and their afterglows at the beginning of the BeppoSAX era. After the increase of the number of detected and localized GRBs, with more and more information on their afterglows and on their spectro-temporal evolution, the situation became more complicated and new developments of the model were needed. These developments will be presented in Chapter 7, after we have explained the standard model in this chapter and the following one.

5.2 SHOCKS AND PHOTON EMISSION

We consider in this section the physical processes which form the basis of our theoretical understanding of GRBs. As explained above, the prompt emission and the afterglow are believed to be due to electrons accelerated to very high energies, which radiate their energy via synchrotron and/or inverse Compton radiation. We briefly describe here the basic physics of the acceleration of electrons in magnetized relativistic shocks, the role of microphysical parameters, and the role of synchrotron and inverse Compton radiation in the generation of GRB spectra. These points will be developed in greater detail in the context of internal shocks in Section 5.3 for the prompt emission, and in Chapter 6 in the context of external shocks for the afterglow.

5.2.1 Relativistic shocks

In the initial phase of the fireball, the relativistic shell coasts at constant $\Gamma = \Gamma_{\text{max}} = \eta$. As it propagates with a Lorentz factor $\Gamma$ it drives a shock into the ISM. This shock propagates with $\Gamma_{\text{sh}} = 2^{1/2} \Gamma$. Behind the shock the ISM is heated. Shocks can also take place within the relativistic ejecta when shells with different Lorentz factors merge. These shocks are collisionless because the densities are so low that the mean free path of the particles for collisions is larger than the typical size of the system. But the ordered or random magnetic fields or plasma waves which are present in the shock region are able to redistribute the energy between the particles.

Shocks are characterized by sharp jumps in the physical parameters. Conservation of mass, energy, and momentum determine the Hugoniot shock jump conditions across relativistic shocks. Blandford and McKee (1976) gave the number density and the energy density $n_{1,2}$ and $e_{1,2}$ measured in the local rest-frame for the matter
upstream (region 1) and downstream (region 2), $\Gamma$ is the Lorentz factor just behind the shock and $\Gamma_{sh}$ is the Lorentz factor of the shock front measured in the rest-frame of the upstream fluid $\Gamma_{sh}^2 = 2\Gamma^2$. Across a relativistic shock, matter is compressed by a factor $\Gamma$. The pressure or the internal energy density behind the shock is of the order of $\Gamma^2 n m_p c^2 = \Gamma^2 \rho_{\text{ext}} c^2$.

In the shock’s rest-frame the relative thermal energy per particle (downstream) is of the same order as the kinetic energy per particle ahead of the shock (upstream). Therefore we can say that the shock converts the ‘ordered’ kinetic energy into a comparable random kinetic energy. If the shock is ultra-relativistic, the downstream random velocities are ultra-relativistic.

Similar jump conditions can be derived for the magnetic fields across the shock. A parallel magnetic field is compressed and amplified by a factor $\Gamma$. Therefore in the downstream region behind the shock the magnetic field is highly amplified and particles can be accelerated to high energies. These are ideal conditions for synchrotron emission to radiate in $\gamma$-rays.

### 5.2.2 Timescales in relativistic flows

This section explains briefly the transformation of timescales between events taking place within the fireball and their detection by an observer at infinity. For a more detailed analysis of these classical phenomena see for instance Mészáros (2006) or Piran (2005). In this section we neglect the effects of the cosmological redshift, which multiplies all times by $(1 + z)$.

The source in the fireball is moving relativistically with a velocity $v$ which is defined by $\beta = v/c = (1 - 1/\Gamma^2)^{1/2}$ (or $\Gamma^2 = 1 - v^2/c^2$), where $\Gamma$ is the Lorentz factor of the relativistically expanding flow relative to the laboratory frame. We consider below two photons emitted by the same electron at radii $R_1$ and $R_2$ along the jet axis. The first photon emitted at $R_1$ will reach the observer before the second photon emitted at $R_2$ (by the same electron) with a time difference $\delta t = [(R_2 - R_1)/v - (R_2 - R_1)/c]$. For $\Gamma \gg 1$, $\Delta T = (R_2 - R_1)/2c\Gamma^2$, or $\Delta T = \Delta R/2c\Gamma^2$ (Figure 5.2, from Sari & Piran 1997b). Under these conditions a characteristic observer timescale is introduced: $T_R \sim R/2c\Gamma^2$. This relation is modified if the expansion velocity is not constant (Sari 1997, Piran 2005). This case is important during the afterglow phase, when the fireball is decelerated in the circum-burst medium.

Another characteristic timescale is usually called $t_{\text{ang}}$. This timescale can be understood in Figure 5.3. This figure shows that the emission along the line of sight to the observer will arrive earlier than the emission from parts of the shell moving at an angle $\theta$ relative to the line of sight. The time difference is $T_{\text{ang}} \sim R/c(1 - \cos \theta)$. As we have noted the visible emission is inside a cone with an opening angle $\theta = 1/\Gamma$, which is very small if $\Gamma$ is large. Thus, $T_{\text{ang}} \sim (R/c)(\theta^2/2) = R/2c\Gamma^2$. Interestingly, this is the same expression as $T_R$. In conclusion, for an observer close to the line of sight the observed time differences $T_R$ and $T_{\text{ang}}$ are characterized by $T \propto R/2c\Gamma^2$. The angular and radial timescales are expressed by the same dependency in $R$ and $\Gamma$. 
Figure 5.2. The various timescales from a relativistic expanding shell with a width $\Delta$ and a Lorentz factor $\Gamma$. They can be defined in terms of the arrival time $t_i$ of various photons: $t_{\Delta} = t_B - t_A$, $t_R = t_C - t_A$, and $t_{\text{ang}} = t_D - t_A$ (from Sari & Piran 1997b). For instance $t_{\text{dur}} > \Delta/c$ with $t_{\text{dur}}$ representing the total duration of the burst. A photon emitted at $A$ (distance $R_1$) will reach the observer $(R_2 - R_1)/2\Gamma c^2$ or $\Delta R/2 \Gamma c^2$ before a photon emitted by the same electron from $C$ at $R_2$. If the emission is continuous from $R_1$ to $R_2$ this characteristic time will limit the duration of the peaks inside a GRB of duration $t_{\text{dur}}$. The same characteristic time is found when $t_{\text{ang}} = t_D - t_A$ is calculated. This is also visible in Figure 5.3.

5.2.3 Microphysical parameters

The synchrotron spectrum can be calculated if we know the detailed physical conditions of the radiating region. Our current understanding of relativistic shocks is not sufficient to allow us to calculate the distribution of magnetic fields and particle momentum in the shocked region. The properties of the shocked region are usually simply described with two dimensionless parameters: $\varepsilon_B$, the ratio of the magnetic energy to the energy dissipated in the shock, $e$, and $\varepsilon_e$, the ratio of the electron energy to the energy dissipated in the shock, $e$.

Clearly the knowledge of $\varepsilon_e$ and $\varepsilon_B$ is critical. For internal shocks, the so-called ‘equipartition’ hypothesis is often used, which assumes that the energy is equally distributed between protons, electrons, and the magnetic field. Moreover, it is
Figure 5.3. This figure illustrates the meaning of $t_{\text{ang}}$ for spherical relativistic shells and how it can be very simply obtained. $\Omega$ is the beaming factor of the GRB emission and $\theta = \frac{1}{\Gamma}$ is the part of this emission which is visible to the observer (from Daigne 1999). This figure shows that $T_{\text{ang}} \sim \Delta R/c \sim R/c(1 - \cos \theta) = R/2c\Gamma^2$.

generally assumed that $\varepsilon_e$ and $\varepsilon_B$ are constant throughout the evolution of the burst. But in fast-evolving flows such as those of GRBs it is not clear whether such equipartition can occur anywhere except perhaps near the shock transition itself. Microphysical parameters strongly depend on the poorly known question of post-shock energy exchange between electrons and protons. The electrons cool quickly compared with protons and it is not trivial to determine whether the protons remain hot behind the shock after the electrons have cooled (i.e. whether a two-temperature plasma is present) or whether they achieve some energy equipartition with the cooled electrons due to unknown fast energy exchange mechanisms. These open questions (Mészáros, Rees, & Wijers 1998) demonstrate the difficulties of fixing $\varepsilon_e$ and $\varepsilon_B$ and knowing their variations through the burst evolution. Of course it is easier to consider $\varepsilon_e$ and $\varepsilon_B$ as constant through the burst.

For external shocks, the microphysical parameters are usually adjusted to reproduce the spectrum and the temporal evolution of the afterglow (e.g. Panaitescu & Kumar 2001). We will come back to this question in Chapter 7, as some authors have suggested that the evolution of the early afterglow could require rapid changes with time of $\varepsilon_e$ and $\varepsilon_B$ (Panaitescu et al. 2006).

5.2.4 Particle acceleration

Another critical question is the problem of particle acceleration in mildly relativistic shocks. This point is very important if we want to estimate the acceleration of
electrons behind the relativistic shock and the resulting electron spectrum. Knowledge of the electron spectrum is needed to calculate the synchrotron spectrum which is at the origin of the GRB and its afterglow.

The common model for particle shock acceleration is the diffuse shock acceleration model: the particles are accelerated when they repeatedly cross a shock. Magnetic irregularities (Alfvén waves) confine the particles for some time near the shock. The formation of a power-law spectrum is the result of the competition between the energy gain per shock crossing cycle (e.g. from upstream to downstream and back) and the probability of escape from the shock per crossing cycle, \( P_{\text{esc}} \). Shock acceleration is a realization of the statistical acceleration process proposed by Fermi (1949). In the context of the acceleration of galactic cosmic rays, diffuse shock acceleration is the mechanism which is proposed to produce energetic particles near strong shocks in magnetized plasmas (Krymskii 1977, Axford, Leer, & Skadron 1978, Bell 1978a,b, Blandford & Ostriker 1978). In fact it is believed that the shocks created by supernova explosions are responsible for the production of Galactic cosmic rays with energy \( E < 10^{15} \) eV.

In the case of GRBs the shocks are relativistic and particle acceleration is more complicated. Diffuse shock acceleration is not applicable because the propagation of accelerated particles near the shock cannot be described as a spatial diffusion. The anisotropies in the angular distribution of the accelerated particles are large and the diffusion approximation for spatial transport does not apply. But the same general principles as in the non-relativistic case operate if the upstream medium contains magnetic fields (and possibly magnetic fluctuations) to deflect or scatter charged particles and if strong magnetic fluctuations are present downstream to allow magnetically scattered particles to return to the shocks (Achterberg et al. 2001). Many authors have discussed particle acceleration in relativistic shocks: Heavens and Drury (1988), Bednarz and Ostrowski (1998), Gallant and Achterberg (1999), Kirk et al. (2000), Achterberg et al. (2001), and Vietri (1995, 2003). It appears that in the first cycle the particles originating upstream have their energy boosted by a factor \( \Gamma_{\text{sh}}^2 \) (\( \Gamma_{\text{sh}} \) is the Lorentz factor of the shock), if they are scattered back across the shock into the upstream region. But for subsequent shock crossing cycles, the particle energy typically doubles (Vietri 1995, Gallant & Achterberg 1999, Achterberg et al. 2001). The energy spectrum of particles accelerated in ultra-relativistic shocks is still a power law, as in the non-relativistic case (e.g. cosmic ray acceleration). The slope of this power law is \( p = 1 + \ln(1/P_{\text{ret}})/\ln(E_f/E_i) \) (with \( N(E) \, dE \propto E^{-p} \, dE \)). Here \( P_{\text{ret}} = 1 - P_{\text{esc}} \) is the mean probability per cycle that a particle returns to the shock and recrosses into the upstream medium. \( E_f/E_i \) is the ratio of final to initial energy in a cycle, the brackets indicating the average values (Achterberg et al. 2001).

A nearly universal value is obtained for the slope of the power law: \( p \sim 2.2–2.3 \). This result has been obtained by several groups with numerical simulations and analytic considerations. This result was, however, contested by Ellison and Double (2002) who indicated that most studies are based on the test particle approximation, which does not include the back reaction of the accelerated particles on the shock structure. They concluded that efficient mildly relativistic shocks do not produce particle spectra close to the ‘universal’ power law. From the observational point
of view, this ‘universal’ slope agrees with the slope inferred from several GRB spectra and afterglow spectra (see for instance Panaitescu & Kumar 2001). Some GRBs, however, do not fit into this figure, with values of $p < 2$ (Panaitescu & Kumar 2001). In this case the analytic model may not be appropriate, since values of $p < 2$ indicate a divergence of the electron energy spectrum (see Section 9.8.2).

Other acceleration mechanisms have been discussed. Rieger and Duffy (2005) proposed the gradual shear acceleration of energetic particles in GRB jets. Hededal et al. (2004) investigated numerically the dynamics of charged particles in highly relativistic collisionless shocks. They found that the power-law distribution of accelerated electrons originates from acceleration processes taking place in power-law distributed filamentary structures resulting from the Weibel-like two-stream instability. Contrary to Fermi acceleration, the electrons are accelerated and decelerated instantaneously and locally. Nevertheless the slope of the resulting power law is in agreement with the observations of GRB afterglows.

5.2.5 Synchrotron emission

Since we have accelerated electrons and amplified magnetic fields, synchrotron radiation of the electrons has to play a role in both the GRBs and their afterglows. We now consider some key features of synchrotron emission as reviewed for instance by Piran (2005). A detailed analysis of the characteristics of this emission can be found in Rybicki and Lightman (1979).

**Characteristics of the synchrotron emission**

Three ingredients are essential for the characterization of the synchrotron emission: $\gamma_e$, the Lorentz factor of the relativistic electrons, $B$, the strength of the magnetic field, and $\Gamma$, the Lorentz factor of the emitting material. Taking into account the fact that the photons seen by the observer are blueshifted because of the relativistic motion of the radiating material, the characteristic synchrotron frequency $\nu_{\text{syn}}$ in the observer frame is:

$$\nu_{\text{syn}}(\gamma_{e, \text{obs}}) = \frac{q_e B}{2\pi m_e c^2 \gamma_e^2 \Gamma}$$

(5.1)

where $q_e$ and $m_e$ are the charge and the mass of the electron.

The power emitted in the fluid co-moving frame by a single electron is:

$$P_{\text{syn}}' = \frac{4}{3} \sigma_T c U_B \gamma_e^2$$

and

$$P_{\text{syn}} = \Gamma^2 P_{\text{syn}}'$$

in the observer frame

where $U_B = B^2 / 8\pi = \varepsilon B e$ is the magnetic energy density, and $\sigma_T$ the Thompson cross-section.

The cooling time of the electron in the fluid co-moving frame is $t' = \gamma_e m_e c^2 / P_{\text{syn}}'$ and $t \sim \Gamma^{-1} t'$ in the observer frame, hence

$$t_{\text{syn}}(\gamma_e) = \left[ \frac{3m_e c^4}{4\sigma_T U_B \gamma_e} \right] 1/\Gamma$$

and

$$t_{\text{syn}} \propto \gamma_e^{-1} \Gamma^{-1}$$
Using equation (5.1) above to replace $\gamma_e$ as a function of the electron frequency in the observer frame, we have:

$$t_{\text{syn}}(\nu) = \frac{3}{\sigma_T} \sqrt{\frac{2\pi m_e c q_e}{B^3 \Gamma}} \nu^{-1/2} \approx \Gamma^{-1/2} \nu^{-1/2}$$

The cooling time at frequency $\nu$ depends on the photon frequency $\nu$, but does not depend on the electron energy $\gamma_e$, as long as electrons with the right $\gamma_e$ are able to produce photons with energy $h\nu$; $t_{\text{syn}}$ gives a lower limit to the variability timescale of GRBs: the spikes in a GRB cannot be shorter than $t_{\text{syn}}$. This is not a very strong constraint, however, since $t_{\text{syn}}$ is much shorter than the duration of GRBs (Piran 2005). Overall, no clear connection has been found between the synchrotron cooling time $t_{\text{syn}}$ and the light-curves of GRBs.

**Synchrotron spectrum**

Now it is important to calculate the overall synchrotron spectrum radiated by a population of electrons with a distribution of energies. Following Sari, Piran, and Narayan (1998), we assume that the energies $\gamma_e$ of the electrons follow a power law, with an index $p$ above a minimal Lorentz factor $\gamma_m$. This is the expected distribution of shock accelerated particles (see Section 5.2.4).

$$N(\gamma_e) \sim \gamma_e^{-p} \quad \text{for } \gamma_e > \gamma_m$$

This leads to $\langle \gamma_e \rangle \propto \gamma_m$.

To keep the energy of the electrons finite, the spectral index $p$ must obey $p > 2$. As most of the electrons have their energy near $E_{\text{min}} = \gamma_m m_e c^2$, this energy can be considered as the characteristic energy of the electron population. Therefore $\gamma_m$ characterizes the ‘typical’ electron’s Lorentz factor and the corresponding ‘typical’ synchrotron frequency: $\nu_m = \nu_{\text{syn}}(\gamma_m)$.

If the electrons are accelerated behind a shock propagating through a uniform and cold medium with density $n$, the particle density and the energy density behind the shock are given by $4\Gamma n$ and $4\Gamma^2 nm_p c^2$ respectively, where $\Gamma$ is the Lorentz factor of the shock fluid (Blandford & McKee 1976). Sari, Piran, and Narayan (1998) assume that a constant fraction $\varepsilon_e$ of the shock energy $e$ goes into electrons then: $\gamma_m = \varepsilon_e (p - 2/p - 1) (m_p/m_e) \Gamma$. They also assume that the magnetic energy density behind the shock is a constant fraction $\varepsilon_B$ of the shock energy, allowing computation of the value of the magnetic field in the fluid frame: $B = (32\pi m_p \varepsilon_p n)^{1/2} \Gamma$. The determination of $P_{\text{syn}}$ in the previous paragraph assumes that the electrons do not lose a significant fraction of their energy to radiation. This is obtained if $\gamma_e < \gamma_c$, where $\gamma_c$ is a critical value of the energy, computed from: $\Gamma \gamma_e m_e c^2 = P(\gamma_e) t$ (where the time $t$ is measured in the frame of the observer, like $P$), hence $\gamma_c = \frac{6\pi m_e c}{\sigma_T \Gamma B^2 t}$. This means that an electron with an initial Lorentz factor $\gamma_e > \gamma_c$ cools down to $\gamma_c$ in a time $t$. As it cools, the frequency of the synchrotron emission varies as $\gamma_c$, while the electron energy varies as $\gamma_e$, and the spectral power $P_\nu \sim P(\gamma_e)/\nu(\gamma_e)$ varies as $\nu^{-1/2}$ over the frequency range $\nu_c < \nu < \nu(\gamma_e)$ ($\nu_c$ is defined by $\nu_c = \nu(\gamma_c)$). The net
spectrum of radiation from such an electron then consists of three segments (Sari, Piran, & Narayan 1998):

- The low-energy part of the synchrotron spectrum is always the sum of the contributions of the tails of the emission of all the electrons: \( F_\nu \propto \nu^{1/3} \). This is a characteristic of synchrotron radiation (Mészáros & Rees 1993a, Katz 1994b, Cohen et al. 1997) and is independent of the exact shape of the electron distribution.

- On the other hand, the most energetic electrons cool rapidly and emit practically all their energy \( \gamma m_e c^2 \) at their synchrotron frequency. The number of electrons with this energy is proportional to \( \gamma^{(1-p)} \) and their energy is proportional to \( \gamma^{(2-p)} \). This energy is deposited in a frequency range \( \sim \nu_{\text{syn}}(\gamma) \propto \gamma^2 \) and therefore \( F_\nu \propto \gamma^{-2} \propto \nu^{-p/2} \). The uppermost part of the spectrum will thus have a power-law dependence, giving a direct access to the index of the electron power-law spectrum.

- In the intermediate frequency region the spectrum varies depending on the cooling regime: slow or fast.

The maximum emissivity appears at \( \nu_c \) and is given by: \( P_{\nu,\text{max}} = (m_e^2 c^2 \sigma_T / 3 q_e) \Gamma B \). To obtain the synchrotron spectrum from a power-law distribution of electrons we have to integrate over \( \gamma \). Two cases have to be considered, \( \gamma_m > \gamma_c \) or \( \gamma_c > \gamma_m \).

**Case 1.** When \( \gamma_m > \gamma_c \) all the electrons cool down to roughly \( \gamma_c \), and the spectral power at \( \nu_c \) is approximately \( N_e P_{\nu,\text{max}} \), where \( N_e \) is the total number of swept-up electrons in the post-shock fluid. This is the case of *fast cooling*. The flux at the observer \( F_\nu \) is given by (Sari, Piran, & Narayan 1998):

\[
F_\nu = \begin{cases} 
(\nu/\nu_c)^{1/3} F_{\nu,\text{max}} & \text{for } \nu_c > \nu \\
(\nu/\nu_c)^{-1/2} F_{\nu,\text{max}} & \text{for } \nu_m > \nu > \nu_c \\
(\nu/\nu_c)^{-1/2} (\nu/\nu_m)^{-p/2} F_{\nu,\text{max}} & \text{for } \nu > \nu_m
\end{cases}
\]

With \( \nu_m = \nu(\gamma_m) \) and where \( F_{\nu,\text{max}} = N_e (P_{\nu,\text{max}}/4\pi D^2) \) is the observed peak flux from a source at a distance \( D \).

**Case 2.** When \( \gamma_c > \gamma_m \) only those electrons with \( \gamma_e > \gamma_c \) can cool. This is the *slow cooling* case because the electrons with \( \gamma_e \sim \gamma_m \) which form the bulk of the population do not cool within a time \( t \). The flux at the observer \( F_\nu \) is given by:

\[
F_\nu = \begin{cases} 
(\nu/\nu_m)^{1/3} F_{\nu,\text{max}} & \text{for } \nu_m > \nu \\
(\nu/\nu_m)^{-(p-1)/2} F_{\nu,\text{max}} & \text{for } \nu_c > \nu > \nu_m \\
(\nu/\nu_c)^{-(p-1)/2} (\nu/\nu_c)^{-p/2} F_{\nu,\text{max}} & \text{for } \nu > \nu_c
\end{cases}
\]

Figure 5.4 from (Sari, Piran, & Narayan 1998) gives the synchrotron spectrum for the two conditions of fast and slow cooling.

Fast cooling must take place during the GRB itself. The relativistic shocks must emit their energy effectively otherwise there would be a problem of inefficiency: the
Figure 5.4. Synchrotron emission from electrons accelerated in a relativistic shock. The electrons have a power-law distribution of energies of index $p$. (a) Fast cooling. This is the situation expected at early times. The scalings above the arrows correspond to an adiabatic evolution and the scalings in square brackets correspond to a fully radiative evolution. (b) Slow cooling (at $t > t_0$). In this case the evolution is always adiabatic. In the two cases there are four segments with three characteristic frequencies $\nu_m$, $\nu_c$, and $\nu_a$. In both cases self-absorption is important below $\nu_a$ (from Sari, Piran, & Narayan 1998).

electrons have to release all their energy rapidly. Moreover if the cooling time were too long the variability would be suppressed. The transition from fast to slow cooling is expected to take place during the early stages of the external shock in the early afterglow phase or late prompt phase (Mészáros & Rees 1997, Waxman 1997a,b, Mészáros, Rees, & Wijers 1998).
Synchrotron self-absorption

At low frequencies (radio) synchrotron self-absorption may take place. It leads to a steep cut-off of the low-energy spectrum, with one of the commonly known dependences: \( \nu^{5/2} \) or \( \nu^2 \) (Rybicki & Lightman 1979). \( \nu^{5/2} \) is obtained when the synchrotron frequency of the electrons emitting the self-absorbed radiation is inside the self-absorption range. \( \nu^2 \) is obtained when the self-absorbed radiation is due to the low-energy tail of electrons that are radiating effectively at higher energy. Synchrotron self-absorption is probably irrelevant during the GRB itself. However, self-absorption appears regularly during the afterglow and is observed typically in radio emission (Katz & Piran 1997, Waxman 1997a, Granot, Piran, & Sari 1999, Wijers & Galama 1999).

5.2.6 Inverse Compton Scattering (ICS)

Even though it is generally accepted that the photons radiated during the prompt GRB phase are due to synchrotron emission, inverse Compton scattering could play an important role during the prompt phase, depending on the exact physical conditions within the fireball. When inverse Compton scattering happens it has a strong impact on the photon. After one single scattering the photon’s energy is so high that it is above the Klein–Nishina energy in the electron rest-frame. The decrease of the Compton cross-section in this energy range makes a second scattering unlikely. The effect of inverse Compton in GRBs has been considered by many authors (Waxman 1997b, Wei & Lu 1998, Panaitescu & Kumar 2000, Dermer, Chiang, & Mitman 2000, Sari & Esin 2001, Björnsson 2001, Wang, Dai, & Lu 2001, Zhang & Mészáros 2001, Wu et al. 2005). Inverse Compton interactions contribute to delay the transition from the early fast-cooling phase to the late slow-cooling phase. When Compton cooling of electrons dominates over synchrotron cooling, the afterglow exhibits quite different behaviors (Li, Dai, & Lu 2002). The impact of Compton cooling on the determination of the basic parameters \( E, n, \varepsilon_e, \varepsilon_B \) has been analyzed by Sari and Esin (2001).

Wu et al. (2005) computed the multi-wavelength afterglow light-curves from a semi-radiative blastwave in the interstellar medium and in a stellar wind, taking into account the important role of IC scattering. The preparation of the GLAST mission has renewed interest in studies of the GRB ICS radiation (see Sections 9.4 and 9.9).

Inverse Compton scattering boosts the energy of the photon by a factor of \( \gamma^2 \) and can introduce an ultra-high energy component in GRB spectra, if \( \gamma \sim 100 \). For typical synchrotron photons in the \( \sim 100\,\text{keV} \) range, the electron IC component will be in the GeV range (Zhang & Mészáros 2001). Such a high-energy component has been observed in some GRBs (see, for instance, GRB 940217 with the detection of GeV photons with Egret (Hurley et al. 1994). Wang, Dai, & Lu (2001), analyzing the IC effect in the very early afterglows, reported that in the TeV energy range the combined IC emission and/or Synchrotron Self Compton emission from the forward shocks becomes dominant for a moderately steep distribution of shocked electrons (\( p \sim 2.5 \)). However, they find that for smaller values of \( p \) (\( p \sim 2.2 \)), the Synchrotron
Self Compton emission from the reverse shock may dominate even at TeV energies. Marginal evidence for TeV emission from GRB 970417A was reported using the Milagrito high-energy detector (Atkins et al. 2000). An excess of gamma-rays above the background was detected coincident both spatially and temporally with the BATSE observation of GRB 970417A. The chance probability of detecting such an excess was estimated to be smaller than \(2.8 \times 10^{-5}\) and the probability for observing an excess at least this large from any of the 54 bursts within the field of view of Milagrito was \(1.5 \times 10^{-3}\), making it difficult to conclude on the reality of the detection. Great advances in this field are expected from AGILE and GLAST (Fermi). These two high-energy satellites will permit accurate measurements of the luminosity emitted by GRBs at gigaelectronvolt energies.

### 5.3 THE PROMPT EMISSION

The prompt GRB (and its afterglow) is due to the dissipation of the kinetic energy of the relativistic flow. This dissipation can take place in an external shock or in internal shocks.

#### 5.3.1 Internal versus external shocks

The external shock model involves the slowdown of the fireball by the external medium. It is the analog of a SNR but with an ultra-relativistic flow. Soon after the introduction of the fireball model, various attempts were made to explain GRBs by the interaction of the fireball with the external medium and the transfer of a substantial part of its kinetic energy to the shocked external medium (Rees & Mészáros 1992, Mészáros & Rees 1993b). If GRBs are the result of an external shock, the burst duration should be related to the deceleration time, the variability due to inhomogeneities in the interstellar medium, and the precursors due to internal shocks within the ejecta (e.g. Sari & Piran 1995). Sari, Narayan, & Piran (1996) returned to the possibility of explaining GRBs by external shocks (the forward and the reverse shock). They explained many properties of GRBs but they had no explanation for the spectral energy breaks seen by BATSE at around a few hundred kiloelectronvolts, and they pointed out some potential problems in explaining the light-curves and the high radiative efficiency of GRBs. Still using external shocks, Panaitescu and Mészáros (1998) showed a remarkable degree of qualitative agreement with a large range of spectral and temporal correlations exhibited by GRBs. They concluded that such models might be responsible for part of the emission of the GRBs or alternatively that a substantial subset of them (i.e. the less variable ones) might be explained by external shock events. The main drawback of external shock models is that they predict smooth GRB light-curves, in contrast to the fast temporal variability (tens of milliseconds) found in many GRBs (e.g. Fenimore, Madras, & Nayakshin 1996, Sari and Piran 1997a,b).
In 1994, Rees and Mészáros, aware of the difficulties faced by the external shock model, proposed the internal shock model. In their paper they stress that internal shocks form at smaller radii and in regions of higher densities compared with external shocks with the ambient matter. The Lorentz factors required to produce gamma-rays are thus lower with internal shocks. This allows slightly larger amounts of baryonic load and leads to less stringent constraints on the source. Later, many authors studied the hydrodynamics of internal shocks (Sari & Piran 1997b, Kobayashi, Piran, & Sari 1997, Daigne & Mochkovitch 1998) and showed that internal shocks offer a natural explanation for the multiple timescales observed in the light-curves of GRBs. Provided that the source itself is variable, the observed temporal structure reflects the activity of the ‘inner engine’ that drives the GRB.

McMahon, Kumar, & Panaitescu (2004), reconsidered the possibility of explaining the prompt γ-ray and early afterglow emissions using the external shock model for some GRBs. Using a sample of 10 GRBs they found that one of them, GRB 970508 (a single-pulse burst), fits nicely with the extrapolation of its afterglow, suggesting that this burst could have been produced by an external shock. Two bursts are possible candidates, provided that the magnetic field energy during the burst is close to equipartition and larger by a factor \( \sim 10^2 \) than the value estimated in the afterglow at \( \sim 1 \) day. The seven remaining GRBs are not compatible with the external shock model. So, for the majority of the sample, internal shocks seem necessary. Ramirez-Ruiz & Fenimore (2000), studying pulse-width evolution in GRBs as a function of the pulse amplitude, found a power law with an index of about 2.8 (their Figure 4). This behavior can be reproduced by the internal shock model adding some quantitative evidence that GRBs are indeed caused by internal shocks.

As the internal shocks can only dissipate a fraction of the kinetic energy, they must be accompanied and followed by external shocks between the flow and the circum-burst medium that dissipate the remaining energy. The external shocks (the forward shock and the reverse shock) produce the afterglow, a smooth long-lasting emission. But they are also invoked for the very early emission: the late part of long GRBs and the prompt optical flashes (see Section 6.6). This is the internal–external shocks scenario. In the next sections, we present this scenario in detail, starting with the internal shock model which is today the favourite scenario for explaining the GRB itself.

### 5.3.2 Internal shocks

This scenario was introduced by Rees and Mészáros (1994), and was briefly mentioned even earlier by Narayan, Paczyński, and Piran (1992). Internal shocks can appear within a variable relativistic wind produced by a highly variable source. If the Lorentz factor of the wind is variable, successive shells can have large relative velocities. A fast shell that is injected after a slower one will eventually catch up and collide with it. This is the origin of internal shocks.

If we call \( \Gamma_s \) and \( \Gamma_f \) the Lorentz factors of the slow and fast shells respectively, and if we assume that \( \Gamma_s \) and \( \Gamma_f \) are of the same order, \( \Gamma \) (within a factor of 2), the time before their collision in the source rest-frame will be \( T = \frac{\Gamma_s^2 \Gamma_f^2}{(\Gamma_s^2 - \Gamma_f^2)} t_{\text{var}}, \) of
order $\Gamma^2 t_{\text{var}}$, where $t_{\text{var}}$ is the variability timescale in the rest-frame. The collision will occur at radius $R = \Gamma^2 t_{\text{var}} c$. For the prompt emission to be detectable, these shocks have to appear when the ejecta have attained sufficiently large distances to allow the resulting radiation to escape without thermalization or adiabatic losses (Rees & Mészáros 1994). For typical values of $t_{\text{var}}$ (1 s) and $\Gamma$ (100), they take place at $R \sim 10^{14}–10^{15}$ cm (see Figure 5.1). Upper limits on $R$ (and so on $\Gamma$) can also be determined because for very high initial Lorentz factors internal shocks appear at distances which are too large and external shocks take place before the internal shocks could appear. Piran (2005) suggested that this fact might play a role in limiting the relevant Lorentz factors and therefore the range of variability of the peak energy $E_p$ of the GRBs.

The total duration of the burst is given by $\Delta$, the total duration of the energy injection (Figure 5.2), and as $\Delta > t_{\text{var}}$ the burst duration is naturally longer than the duration of the peaks (and it can be much longer for long, complex GRBs). Moreover, in the internal shock model, the light-curve of the prompt emission follows the activity of the central engine. This permits production of the complex temporal structures observed in GRB light-curves. In summary, in the internal shock model the (unknown) continuous and variable inner engine has to operate during the whole duration of the burst and to vary on the observed variability timescale.

To examine the energy conversion in internal shocks various models have been considered. Kobayashi, Piran, and Sari (1997) and Daigne and Mochkovitch (1998) represent the ultra-relativistic wind by many shells with different Lorentz factors and densities. In these models the collision of two shells is the elementary process allowing the acceleration of electrons. Daigne and Mochkovitch (1998), for instance, consider a succession of layers emitted every 2 ms with a varying Lorentz factor; since the mass of the layers is proportional to $1/\Gamma$ the energy injection rate is constant. They add a random fluctuation to the Lorentz factor of each of the layers injected in the wind in order to reproduce the rapid temporal variations seen in most observed bursts. The short-term variability introduced in the Lorentz factors is reproduced in the GRB profile (Figure 5.5; Daigne & Mochkovitch 1998).

The Lorentz factor of the merged shell $\Gamma_m$ is calculated by:

$$\Gamma_m = [(m_f \Gamma_f + m_s \Gamma_s)/(m_f/\Gamma_f + m_s/\Gamma_s)]^{1/2}$$

The fast shell is characterized by its mass $m_f$ and Lorentz factor $\Gamma_f$ and the slower one by $m_s$ and $\Gamma_s < \Gamma_f$. The internal energy $E_{\text{int}}$ of the merged shell is the difference of kinetic energy before and after the collision:

$$E_{\text{int}} = m_f c^2 (\Gamma_f - \Gamma_m) + m_s c^2 (\Gamma_s - \Gamma_m)$$

In a simple collision the efficiency of conversion of the kinetic energy into internal energy that will be partly emitted as radiation is:

$$\varepsilon = 1 - (m_f + m_s) \Gamma_m/(m_f \Gamma_f + m_s \Gamma_s)$$

The interaction of the two shells takes place in the form of two shocks, a forward and a reverse shock, and their Lorentz factors can be calculated (Sari & Piran 1995).
With some assumptions about the energy equipartition between the electrons, protons, and magnetic field, the Lorentz factors of the accelerated electrons can be determined and their synchrotron emission can be calculated. A fraction of the synchrotron photons can be shifted to higher energy by inverse Compton scattering.

The instantaneous synchrotron spectrum that is obtained is also valid during the subsequent afterglow phase (see Section 6.3). However, the prompt GRB phase involves simultaneous emissions from multiple shocks and the combined spectrum might be more complicated. This can explain why the afterglow is well fitted with a synchrotron spectrum while the situation is not so clear for the prompt GRB spectra (see the discussion at the end of the section).

5.3.3 GRB light-curves and internal shocks

Superimposing the pulses resulting from individual two-shell interactions, synthetic bursts can be constructed (see for instance Figure 5.5, from Daigne & Mochkovitch 1998). After some time, the different shells are organized with the fast shells moving ahead and the slower ones behind, and two-shell interactions stop. This simple model was confirmed by more elaborate hydrodynamical calculations using a 1D
Lagrangian hydrocode to follow the evolution of the relativistic wind (Daigne & Mochkovitch 2000).

These models can produce the highly variable profiles observed in most GRBs and they show that the observed temporal structure reflects the activity of the source. Playing with the initial distribution of Lorentz factors it is possible to reproduce the large variety of observed GRB light-curves. The separation between the peaks corresponds to periods of time during which the inner engine is quiet (Kobayashi, Piran, & Sari 1997). Spada, Panaitescu, & Mészáros (2000) developed another model producing internal shocks in a transient unstable relativistic wind. They compared the temporal properties of their simulated bursts with real GRBs. They identified a set of parameters of the relativistic wind leading to bursts whose temporal properties are very much like those of real GRBs: their average power density spectra exhibit an $f^{-5/3}$ behavior (Beloborodov, Stern, & Svensson 1998) and the time intervals between peaks and the pulse fluences follow log-normal distributions (McBreen et al. 1994, Li & Fenimore 1996). In addition, Nakar and Piran 2002 developed an analytic model to clarify the relation between the observed gamma-ray light-curve and the activity of the inner engine. They explained why the observed light-curves reproduce the activity of the inner engine. They also considered the timescales that determine the pulse width in the burst: the angular time $t_{\text{ang}}$ which results from the spherical geometry of the shells ($t_{\text{ang}} \sim R/2\Gamma_{\text{sh}}^2$), the hydrodynamic time $t_{\text{dyd}}$, which arises from the shell’s width, and the shock crossing time and the cooling time $t_{\text{c}}$ of the emitting electrons. The authors assumed that this last time is shorter than the other two and they showed that in this case the pulse width $\delta t$ is dominated by $t_{\text{ang}}$. Since this time depends on $\Gamma_{\text{sh}}^2$ and as $\Gamma_{\text{sh}}$ has a strong dependence on the ratio of the shell masses, they conclude that the pulse width $\delta t$ will depend strongly on the shell’s masses. But all the observed properties (spectral and temporal) are not necessarily reproduced by these models.

Daigne and Mochkovitch (2003) found that the observed spectro-temporal evolution can only be reproduced when special values of the parameters $\varepsilon_e$ and $\varepsilon_B$ are considered. These authors also revisited the efficiency $\alpha$ and the microphysical parameters $\varepsilon_e$ and $\varepsilon_B$ in the internal shock model, considering a relativistic wind where a slow shell of mass $M_s$ and Lorentz factor $\Gamma_s$ decelerates a more rapid part of the flow characterized by a constant mass $M_f$ or constant energy flux $E$ (in the source frame) and Lorentz factor $\Gamma_f > \Gamma_s$. The flow evolution is obtained by considering that fast material is ‘accreted’ by the slow shell. The increase of its Lorentz factor can be calculated as a function of the accretion rate and, finally, versus time. The dissipated power is obtained and assuming this energy is radiated by the synchrotron process temporal profiles of synthetic pulses can be derived. The standard assumption to estimate the post-shock magnetic field and the electron Lorentz factor, the assumption of equipartition, gives a spectral evolution of $E_p$ (the peak energy) that is too steep. To have a better general agreement with observations the authors assume that the equipartition parameters depend on the strength of the shock and on the post-shock density. Under these conditions the hardness/intensity and the hardness/fluence correlations as well as the time lags between profiles in different energy channels can be reproduced. The authors conclude that, if GRB pulses are produced by
internal shocks, their temporal and spectral properties are probably governed by the hydrodynamics of the flow rather than by the geometry of the emitting shells (see also Fenimore, Madras, & Nayakshin 1996, Kobayashi, Piran, & Sari 1997). Such a distinction is based on the comparison between $t_{\text{dyn}}$, the dynamical time taken by internal shocks to travel throughout the flow, and $t_{\text{ang}}$, the angular spreading time, which is the delay in the arrival times of photons emitted by a spherical shell. Daigne and Mochkovitch (2003) noted that, if the thickness of colliding shells is small compared to their initial separation ($t_{\text{dyn}} \ll t_{\text{ang}}$), the pulse temporal evolution is fixed by the geometry. But if the source produces a continuous wind rather than a series of discrete shells, then $t_{\text{dyn}} > t_{\text{ang}}$ and hydrodynamical effects control the pulse shape.

Beyond these discussions on the relative importance of $t_{\text{ang}}$ and $t_{\text{dyn}}$, there is a general agreement about the fact that internal shocks are able to account for the properties of the GRB light-curves. Nevertheless the prompt emission of GRBs still raises important questions. We discuss two of them below: the compatibility of the spectral shape of the prompt emission spectra with the prediction of the synchrotron-shock model, and the radiative efficiency of the internal shock model.

### 5.3.4 Prompt GRB spectra and the synchrotron shock model

One important open issue concerns the compatibility of GRB spectra with the optically thin synchrotron emission from isotropically distributed shock-accelerated electrons. While the synchrotron shock model has demonstrated its success for many bursts (see, for instance, the fits of GRB 910601, GRB 910814, and GRB 920622 by Tavani 1996), some GRBs have a low-frequency spectral slope which seems too hard to be explained by this model.

GRB spectra are characterized by $E_{\text{peak}}$, the energy of the peak of the $\nu F_{\nu}$ spectrum, and two power laws extending below and above $E_{\text{peak}}$. The indices of the photon spectrum are usually called $\alpha$ below $E_{\text{peak}}$, $(N(E) \propto E^\alpha)$, and $\beta$ at high energies (see Band et al. (1993) and Chapter 2). Katz (1994b) discussed the expected slope, $\alpha$, of the low-frequency spectra of GRBs.\footnote{In the comparison between observed spectra and the models, either $F_{\nu} \sim \nu^s$ or $N(E) \sim E^\alpha$ are used. The following relation holds between $s$ and $\alpha$: $\alpha = s - 1$; for instance $s = 1/3$ means $\alpha = -2/3$.} Optically thin relativistic synchrotron radiation leads to a general prediction: $\alpha \leq -2/3$. In fact $\alpha = -2/3$ is the instantaneous synchrotron spectrum of electrons accelerated by a relativistic shock in one region, if $\gamma_e$ stays constant. But as noted by Katz (1994b) and by Sari and Piran (1997b), the radiating electrons lose most of their energy through synchrotron radiation in a time much shorter than the duration of a typical burst. Therefore, for comparison with spectra integrated over the duration of a GRB, the theoretical spectrum should be integrated over the spectrum emitted by a single electron as its energy decays. As we have seen in Section 5.2.5, $F_{\nu}$ is characterized by a frequency dependency $F_{\nu} \sim \nu^{-0.5}$, leading to $\alpha = -1.5$. So there is a significant difference between the instantaneous spectrum ($\alpha = -2/3$) and the spectrum integrated over the radiative decay of the electron energies ($\alpha = -3/2$; see for instance Ghisellini,
Celotti, & Mazzati 2000). The value $\alpha = -2/3$ has been called the ‘line of death’ of synchrotron radiation models.

5.3.5 The ‘line of death’ of synchrotron emission models

Cohen et al. (1997) examined 11 bursts from the Large Area Detectors on BATSE and reported that, below the characteristic synchrotron frequency (the peak of $\nu F_\nu$), all the observed asymptotic spectral slopes lie in the predicted range $-3/2 \leq \alpha \leq -2/3$. Thus, GRBs exhibit asymptotic low-energy spectral indices which correspond to all possible cases between negligible radiation losses of the electrons and instantaneous radiation of all the electron energy. This led the authors to conclude that the bursts with $\alpha = -2/3$ should have long radiation times and therefore smooth profiles while bursts with $\alpha = -3/2$ should have short radiation times and therefore may have spiky, multi-peaked profiles. A simple interpretation of their data concerning this correlation between spectrum and time structure would be that the electron radiation time is comparable to the width of a spike or subpeak in the burst.

Other authors did not share the same optimism, noting that several GRBs have values of $\alpha$ greater than $-2/3$ (e.g. Crider et al. 1997). Considering a much larger sample of GRBs detected by BATSE, Preece et al. (1998) found 23 bursts out of 137 bursts in which the ‘line of death’ ($\alpha > -2/3$) is violated. Ghirlanda et al. (2003) also found some BeppoSAX GRBs with very hard spectra at low energies, clearly incompatible with the predictions of synchrotron models (Figure 5.6). The basic synchrotron shock model seems to be in contradiction with such GRBs, and the authors looked for some solutions to accommodate this observation. The first idea was to note that Compton upscattering of soft photons by an energetic distribution of particles can significantly modify the basic synchrotron emission spectrum since energetic particles boost their own synchrotron emission into the observed gamma-ray band (Liang et al. 1997). Another mechanism which tends to increase

![Figure 5.6. Low-energy power-law spectral index of time-resolved spectra of bright bursts detected with BATSE and BeppoSAX. The spectrum chosen for each burst is the one with the largest value of $\alpha$. The two lines of death are given corresponding to $\alpha = -2/3$ and $-3/2$, for $N(E) \propto E^\alpha$ (see text, from Ghirlanda, Celotti, & Ghisellini 2003).](image-url)
the low-energy spectral index $\alpha$, is synchrotron self-absorption; the maximum photon spectral index would be $3/2$ and all the bursts in the study are consistent with this index (Figure 5.6; Preece et al. 1998, Panaitescu & Mészáros 2000).

In 2000, Ghisellini et al. returned to this discrepancy between the synchrotron shock model and the observations. They emphasized the fact that typical spectra due to synchrotron emission in the standard internal shock model should have a slope $F_\nu \propto \nu^{-1/2}$ ($\alpha = -3/2$) because, as we have said, the integration and dynamical timescales are much longer than the particle cooling timescales. Such a prediction strongly conflicts with the much harder observed spectra. This limit dramatically increases the discrepancy between the predictions of the standard model and the observations. Ghisellini, Celotti, and Lazzati (2000), facing this difficulty and without excluding the standard model, tried other alternatives to eliminate this major problem: they considered deviation from equipartition, particle re-acceleration, strongly varying magnetic fields, and adiabatic losses, and concluded that none of these alternatives was adequate to remove the discrepancy. Introducing inverse Compton scattering at a level comparable to the synchrotron process does not solve the problem either. Quasi-thermal comptonization in the dense photon environment of internal shocks implying the presence of mildly or sub-relativistic electrons due to shell–shell collisions has been proposed by various authors as a viable alternative radiation mechanism (Thompson 1994, Liang 1997, Stern 1999, Liang et al. 1999, Ghisellini & Celotti 1999). The self-absorbed synchrotron spectrum produces the soft seed photons for the comptonization process. The resulting spectrum is flat, with $F_\nu \sim \text{cst} \ (\alpha = 0)$. Lloyd and Petrosian (2000) studied the distribution of $\alpha$ in detail. They found that $\alpha$ is generally above $-1.5$, but they did not consider this discrepancy as catastrophic. They believed instead that spectral indices greater than $\alpha = -2/3$ can be obtained if one includes the possibility of synchrotron self-absorption, if one relaxes the assumption of a sharp low-energy cut-off in the spectrum of the accelerated electrons and/or if one assumes small pitch angles for the radiating electrons. Mészáros & Rees (2000) reviewed the work on this question in a paper in which they reexamined the relative roles of photospheres and shocks, as well as those of synchrotron emission, pair breakdown, scattering on MHD waves and comptonization. Their main conclusion is that a better agreement of the models with the observations could be obtained on condition that a larger role is allowed for photospheres and comptonized pairs.

Another solution, suggested by Medvedev (2000), involved synchrotron radiation in highly non-uniform short-scale magnetic fields. In these conditions, Medvedev showed that the radiation from accelerated electrons does not have the classical synchrotron spectrum and he called this radiation ‘jitter’ radiation. This jitter radiation also has a ‘line of death’, but now corresponding to $\alpha = 0$, i.e. the line of death is shifted to harder spectra, becoming compatible with the great majority of observed GRBs. For instance, only two bursts out of those studied by Preece et al. (1998) are above the $\alpha = 0$ line.

Preece et al. (2002) considered GRBs with low-energy power-law spectral indices $\alpha < -2/3$ and made additional tests of the synchrotron shock model. Specifically, they compared the low- and high-energy spectral indices, assuming that the same
power-law distribution of electrons is responsible for both the low- and high-energy part of the spectra. They found an inconsistency between the observed spectral indices and the relationships predicted by the simple, optically thin, synchrotron shock emission model. They concluded that the synchrotron shock model during the prompt phase is irreconcilable with observations. They also noted that most spectra violate the limit at \( \alpha = -3/2 \), implying that particle acceleration during the prompt emission should not be ignored. In the same paper, Preece et al. (2002) raised another point: in the simple version of the synchrotron model, the high-energy index \( \beta \) is related to the energy spectral index of the accelerated electrons. Their analysis yields \( p \sim 2.8 \), while \( p \sim 2.25–2.3 \) is expected from Fermi acceleration theory. This poor agreement allows the authors to question whether Fermi acceleration is really the mechanism for particle acceleration in GRBs. Ghirlanda, Celotti, and Ghisellini (2002, 2003) studied a number of very bright GRBs, suitable for time-resolved spectral analysis, and found that the low-energy part of the spectrum can be extremely hard at the very beginning of some bursts, with \( \alpha \) in the range +0.5 to +2. Their study of various spectral models led them to conclude that these spectra could be explained by a black body at several tens of kiloelectronvolts (see also Blinnikov, Kozyryeva, & Panchenko 1999).

In 2006 Derishev returned to the issue of the synchrotron emission in the fast-cooling regime. He discussed in particular the question of the low-frequency spectral index, and the fact that the observed spectra appear in general too soft. The expected slope \( F(\nu) \sim \nu^{-1/2} \) is obtained if the synchrotron efficiency is constant. Derishev (2006) noted that this is not the case when a consistent Synchrotron Self Compton model is considered, with prevalence of inverse Compton radiative losses. In this case, the synchrotron efficiency \( \eta \) can rise as fast as \( \eta(\gamma) \sim \gamma \) where \( \gamma \) is the Lorentz factor of the radiating electron. The low-frequency spectrum of synchrotron radiation obtained under these conditions can be rather hard, with \( F(\nu) \sim \text{cst} \), but this is at the price of a decreased synchrotron efficiency and this solution is difficult to apply when the comptonizing electrons are in the Thompson regime. In the same paper, Derishev proposed another way of obtaining hard low-frequency spectra: he noted that the magnetic field behind the shock front, produced by plasma instabilities, is not frozen in and must decay as the shocked plasma moves away from the shock front. Under these conditions, the various parts of the electron distribution function are subject to different magnetic strengths (see his Figure 4). With a strength of the magnetic field which is a power-law function of the distance from the shock front, \( B \propto r^{-\gamma} \), Derishev found that \( F(\nu) \sim \nu^{(-1+\gamma)/(2+\gamma)} \). With this process, spectra as hard as \( F(\nu) \sim \nu \) (\( s = 1, \alpha = 2 \)) can be obtained (but the low-frequency spectrum remains softer than \( F(\nu) \sim \text{cst} \), unless \( \gamma > 1 \)). In summary, hard spectra (e.g. \( F(\nu) \sim \text{cst} \)) are possible for the Synchrotron Self Compton model with comptonization in the Klein–Nishina domain, and even harder spectra are possible if the magnetic field decays quickly behind the shock. This work demonstrates that the question of the GRB line-of-death could be less critical for synchrotron models than was assumed in early analyses.

Clearly, the explanation of the prompt emission of GRBs by the synchrotron shock model is not straightforward. We will see in Chapter 6 that it is much more
successful in explaining the afterglows. The fundamental difference between the two phases can explain such a result: the GRB phase is characterized by an input of energy, manifested in particle acceleration which has to be taken into account, while the afterglow is powered solely by the stored kinetic energy of the blast wave (Preece et al. 2002).

The nature of the emission process(es) leading to GRB prompt emission remains a major issue. It is possible that several processes are at work simultaneously in the formation of GRB emission spectra. At this point a word of caution should be given about the measurement of the low-energy spectral index $\alpha$. The energy range over which $\alpha$ can be measured is in general quite small because the low-energy threshold of GRB detectors is usually around $20$ keV while $E_{\text{peak}}$ takes on values between $100$ keV and $1$ MeV. In addition, the energy response of gamma-ray detectors close to their low-energy threshold is difficult to measure accurately. The combination of these two effects makes the measurement of $\alpha$ for a large number of GRBs a challenging task. The best measurements can be obtained with an instrument having a low-energy threshold at a few kiloelectronvolts. This was the case of BeppoSAX (see Chapter 3) and HETE-2 (see Chapter 4) whose energy range extended down to $2$ keV. Frontera et al. (2000) studied the spectra of eight GRBs and found that the optically thin, synchrotron shock model provides an acceptable representation of most of the time-resolved GRB spectra extending down to $2$ keV, except in the initial phases of several bursts. One event, the quite strong GRB 970111, was not fitted adequately by the synchrotron shock model, showing a clear depletion of photons at low energy. Barraud et al. (2004) measured $\alpha$ for 42 GRBs detected by HETE-2, and found all of them in the range $-3/2 < \alpha < -2/3$. These results are probably not in contradiction with the observations of BATSE and BeppoSAX since they were obtained for time-integrated spectra, which do not cross the ‘line of death’ as often as the time-resolved spectra.

Another question was raised by Soderberg and Fenimore (2001) concerning the signature of curvature from the emitting shell as the result of spherical kinematics. They noted that the observed pulses do not evolve in the manner that curvature demands; instead a sizable fraction of observed pulses evolves faster than kinematics allows. This question was also addressed by Daigne and Mochnovitch (2003). These authors simulated GRB pulses produced in internal shocks and studied their temporal and spectral properties. They concluded that the properties of GRB pulses are probably governed by the hydrodynamics of the flow rather than by the geometry of the emitting shells, removing the objection raised by Soderberg and Fenimore (2001).

### 5.3.6 Efficiency of the internal shock model

The theoretical efficiency of internal shocks to convert the wind kinetic energy into hard X-ray photons is low. Kobayashi, Piran, and Sari (1997) found an efficiency of only a few percent if the spread in the Lorentz factors is relatively low (a factor of 2–3). An efficiency of 10% can be reached easily if the spread of Lorentz factors is higher (~6) and the number of shells is high ($N \sim 100$). This agrees with the results of
Mochkovitch, Maitia, & Marques (1995) and Daigne and Mochkovitch (1998, 2000) who also find an efficiency smaller than 10%. This low efficiency of internal shocks has also been noted by Kumar (1999) who argues that the conversion efficiency from bulk motion to gamma-ray radiation cannot be higher than 1% (see also Panaitescu, Spada, & Mészáros 1999, Spada, Panaitescu, & Mészáros 2000). Very low radiative efficiencies (a few percent) have dramatic consequences: as 50% of the energy of the initial explosion could be lost to neutrinos during the initial phase of the burst, the total energy budget of the brightest GRBs might reach $10^{54}-10^{55}$ erg. Even if the anisotropy of the explosion could reduce this energy requirement by a factor of 10–100 it leaves $\sim 10^{52}-10^{53}$ erg to be found. This is a huge amount of energy, one or two orders of magnitude larger than the kinetic energy of typical supernovae.

The low efficiency of internal shocks can be checked directly for few GRBs that have radio follow-up extending over several months. Such observations permit the ‘calorimetry’ of the burst: the measurement of the total kinetic energy deposited by the GRB in its environment independently of the beaming angle of the ejecta. Three bursts have been observed in sufficient detail to measure their kinetic energy: GRB 970508 and GRB 980703, which have kinetic energies of the order of $3 \times 10^{51}$ erg (Berger, Kulkarni, & Frail 2004), and GRB 030329 with a kinetic energy of $\sim 10^{51}$ erg (Berger, Kulkarni, & Frail 2003). Comparing these values with the beaming-corrected energy radiated by these bursts in $\gamma$-rays gives the radiative efficiency of the internal shocks, respectively 3%, 20%, and 6% (Berger, Kulkarni, & Frail 2004). Overall these numbers seem to be in good agreement with the predictions of the models, suggesting that energetic GRBs (those with $E_{\text{iso}} \sim 10^{53}$ erg) are associated with very narrow jets (with opening angles of a few degrees). Nevertheless, given the importance of this issue and the fact that calorimetry could be done for only three GRBs, various authors have investigated ways to increase the radiative efficiency of internal shocks. There is another reason to work in this direction: if internal shocks have a low efficiency a large fraction of the energy is available for the afterglow phase, possibly even more than in the prompt GRB phase. This appears to contradict the fact that during the afterglow only a tenth of the GRB energy is observed (Frontera et al. 2000, Kumar & Piran 2000a).

Higher radiative efficiencies can be obtained if the spread in $\Gamma$ is larger. For example 10% is obtained for $\Gamma_{\text{max}}/\Gamma_{\text{min}} = 6$ for a number of shells $N = 100$ and with shells of equal energy content (Kobayashi, Piran, & Sari 1997). Beloborodov (2000) and Kobayashi and Sari (2001) proposed another solution to increase this efficiency. They showed that if the distribution of Lorentz factors is logarithmically uniform the typical ratio of Lorentz factors between neighboring shells is considerably larger than predicted by all previous models. Ultra-efficient internal shocks are obtained if colliding shells that do not emit all their internal energy are reflected off each other, causing subsequent collisions and thereby allowing more energy to be radiated. With this scenario all the internal energy is not emitted in each collision, and this is more realistic because the electrons do not carry all the internal energy; a large amount of it remains in protons. This remaining internal energy is transformed back to kinetic energy. Therefore the number of collisions between the shells is greatly increased allowing more energy to go into the electrons even if each collision has a low
efficiency. Kobayashi and Sari (2001) give an example where they obtain an overall efficiency of 60% provided that the shell’s Lorentz factor varies between 10 and $10^3$, even if the fraction of the energy going into the electrons is $\varepsilon_e \sim 0.1$. In a similar approach Belodorov (2000) noted that the fraction of the fireball kinetic energy that is radiated by internal shocks is sensitive to the amplitude of the initial fluctuations in the fireball: with large amplitudes the dissipation occurs in a nonlinear regime with an efficiency approaching 100%, leaving only a small fraction of energy for the afterglow.

Guetta, Spada, and Waxman (2001), gave radiative efficiencies that do not agree with the very high values obtained by Kobayashi and Sari (2001) and Beloborodov (2000). They think that the range of Lorentz factors considered by these authors (from 10 to $10^4$) is too large, because in their opinion the minimum wind Lorentz factor must be significantly higher than 10 (otherwise the wind becomes optically thick) and the maximum Lorentz factor must not exceed $10^3$ due to acceleration process limitations. With their model Guetta, Spada, and Waxman (2001) found that the fraction of the wind kinetic energy which can be converted into radiation can only reach 15%, provided the distribution of the Lorentz factors within the wind has a large variance (for instance with a bimodal distribution of shell Lorentz factors; see their Figure 1) and the minimum Lorentz factor is greater than $\Gamma \times 10^{2.5} L_{52}^{2/9}$ where $L_{52}$ is the wind luminosity in units of $10^{52}$ erg s$^{-1}$.

So even though solutions to increase internal shocks efficiency have been proposed this is still a critical issue. While beaming with an opening angle of a few degrees (Kulkarni et al. 1999, Harrison et al. 1999, Sari, Piran, & Halpern 1999) reduces energy estimates by a factor of 100, a relatively low efficiency of conversion of kinetic energy to $\gamma$-rays might require fireballs carrying a kinetic energy larger than $10^{53}$ erg. According to the internal–external shock model a comparable amount of energy should be released during the GRB phase and the afterglow. However, in general, as we have just said, only a fraction of the energy emitted as $\gamma$-rays is observed in the afterglow, mostly as X-rays, and there is no or limited correlation between $\gamma$-ray fluence and the afterglow flux.

One way to progress in this domain would be to measure the afterglow kinetic energy for a large number of GRBs (and not just for very few bursts with extensive radio follow-up, as discussed above). Kumar (2000), Piran et al. (2001), and Freedman and Waxman (2001) proposed using the X-ray flux measured 10 h after the burst to infer the kinetic energy of the ejecta in the adiabatic afterglow phase. At that time the average Lorentz factor of the ejecta has decreased to $\Gamma \sim 10$, and the inhomogeneities at the surface of the ejecta are smoothed out. In addition, the cooling frequency is below the X-ray range, the afterglow is in the ‘fast cooling’ regime and its luminosity is mostly independent of the surrounding medium. Piran et al. (2001) argued that the kinetic energy measured in this way is within a factor of 2 of the initial energy in the relativistic ejecta. From the narrowness of the X-ray flux distribution, they concluded that the energy of the relativistic ejecta spans less than one order of magnitude. Berger et al. (2003) performed a more complete analysis on 41 GRB afterglows and reached the conclusion that the X-ray luminosity after a few hours, corrected for beaming, has a very small dispersion, within a factor of 2. But unlike
Piran et al. (2001) it is not the isotropic–equivalent X-ray luminosity normalized to $t = 10 h^{-1} L_{\text{iso}}$, which is narrowly distributed, but $L_x$, the true X-ray luminosity: $L_x = f_b^{-1} L_{\text{iso}}$ where $f_b^{-1}$ is the beaming factor. Their Figure 1d illustrates quite well this strong clustering (see also Figure 6.14). It implies that the adiabatic blastwave kinetic energy in the afterglow phase is tightly clustered. Again this suggests that the energy imparted by the central engine to the relativistic ejecta is approximately constant. While these studies provide convincing evidence that GRBs are tapping a nearly constant energy reservoir, they bring no direct insight into the true size of this reservoir and do not constrain the radiative efficiency of the ejecta during the prompt GRB phase.

Kumar and Piran (2000a) proposed a solution to this potential efficiency problem, by assuming the existence of large angular inhomogeneities within the relativistic ejecta. Even if the ejecta are confined within a beam with an angular width of several degrees, the relativistic beaming (the region visible to a distant observer) is $\sim \Gamma^{-1}$, and when $\Gamma$ is high (several hundred), the angular size of the visible region is quite small ($< 0.01 \text{ rad}$). Thus the observed $\gamma$-ray luminosity seen by different observers for the same burst could fluctuate strongly if there are small-scale inhomogeneities in the emitting regions. In this context the detection of a strong burst will correspond to the observation of a hot spot. Later on, when the Lorentz factor of the afterglow has substantially decreased and a large fraction of the emitting region becomes visible, the dispersion of afterglow luminosity seen along different lines of sight will be largely reduced. If this concept is correct the afterglow emission would be a better indicator of the overall total energy involved in a GRB than the prompt emission. To support this model Kumar and Piran (2000a) argued that if the observed large dispersion of the $\gamma$-ray energies ($E_{\text{iso}}$) were due to a broad range of explosion energies, the distribution of X-ray afterglow luminosities should be very broad. As we have just seen this is not the case; the distribution of observed X-ray afterglow luminosities is significantly narrower (by 1.5 dex) than the distribution of $E_{\text{iso}}$. This suggests that the distribution of the total energy is narrow and that the large dispersion of $E_{\text{iso}}$ is due to the variable efficiency of the conversion of the total energy into highly relativistic ejecta. In fact, despite the broad observed luminosity function of GRBs, the total energy in GRB explosions could be quite comparable from one burst to the other. The result is that the gamma-ray fluence is not a good measure of the total energy released in the explosion.

The discussions about the radiative efficiency of internal shocks have been revived by the discovery by Swift of a plateau phase in the early X-ray afterglow of many GRBs (Nousek et al. 2006). This plateau has been attributed either to refreshed shocks or to a long-lasting activity of the central engine. These observations and their consequences are discussed in Chapter 7.

To conclude, it should be mentioned that despite the open questions related to the role of shock synchrotron emission and to the radiative efficiency of internal shocks, the internal shock scenario remains the favorite explanation for the prompt emission of GRBs. However, given the complexity of the issues involved in the production of the prompt GRB emission and of the many uncertainties remaining, other models have been proposed, which are briefly discussed below.
5.3.7 Producing the prompt GRB emission without internal shocks

Poynting flux–dominated models

Magnetic fields are also a natural way to transmit the energy from the central object with a small contribution of baryons. In this case the Poynting flux is involved. In ideal magneto-hydrodynamics (MHD) it can be considered as magnetic energy carried by the outflow from the central object. It is defined by $\mathbf{S} = c\mathbf{E} \times \mathbf{B}/4\pi$. In ideal MHD the electric field is: $\mathbf{E} = \mathbf{v} \times \mathbf{B}/c$ where $\mathbf{v}$ is the fluid velocity. The Poynting flux is thus $\mathbf{S} = \mathbf{v}_\perp (B^2/4\pi)$; $\mathbf{v}_\perp$ being the velocity component perpendicular to the field lines. This clearly indicates that the Poynting flux represents magnetic energy carried by the flow. But it is not just the magnetic energy density $B^2/8\pi$, which accounts for only half of the Poynting flux. The other half can be accounted for by interpreting $\mathbf{S}$ as the flux of magnetic enthalpy $w_m = U_m + P_m$ where $U_m$ and $P_m$ are the magnetic energy density and the magnetic pressure, both equal to $B^2/8\pi$. This distinction is crucial when the acceleration of the flow is considered (Spruit & Drenkhahn 2004). Their analysis of the magnetically powered prompt emission of a GRB and the outflow acceleration is particularly interesting and useful to clarify these magnetic processes, which might power GRBs. We use their analysis here.

The Poynting flux being defined, let us see how the magnetic energy is dissipated. Through instabilities and reconnections, the magnetic energy in the flow can be converted into kinetic energy of the plasma or into fast particles and from there into heat or radiation. Depending on the speed of this energy dissipation it can take place close to the central object or outside the photosphere, if the dissipation is slow. In this last case the environment can be optically thin and the energy of the fast particles produced by reconnection can be radiated as synchrotron emission in the presence of the outflow magnetic field. At the same time the dissipation of magnetic energy reduces the total pressure, and a pressure gradient appears that accelerates the flow outward. Hence, as indicated by Drenkhahn & Spruit (2004), a magnetic central engine can provide both the acceleration of the flow and the dissipation outside the photosphere needed for efficient prompt radiation, the dissipation of the magnetic energy in the flow being due to magnetic reconnections. Moreover as the magnetic fields outside the photosphere are large enough ($10^7–10^8$ gauss) synchrotron emission can take place with high efficiency. So the crucial ingredients for prompt gamma-ray emission are the natural outcomes of magnetic dissipation in a magnetically powered outflow.

The authors also note a very interesting point. The acceleration of the outflow is proportional to the rate of energy dissipation and the kinetic energy attained by the flow equals the magnetic energy dissipated. Hence the acceleration of the flow is possible by destroying the ingredient that drives it. This strange situation can be explained because the energy equation contains the Poynting flux $\mathbf{v}_\perp (B^2/4\pi)$, i.e. twice the flux of magnetic energy. So the acceleration of the flow ‘uses’ half of the Poynting flux. The other half of the available energy through the Poynting flux can be used as thermal energy for the particles or radiation. This clearly illustrates the distinction which has to be made between the flux of magnetic energy, $c(B^2/8\pi)$, and the Poynting flux, $c(B^2/4\pi)$. This quite short analysis of a part of the Drenkhahn
and Spruit review allows us to show that models involving magnetic dissipation are quite attractive for at least three reasons:

- Since the magnetic energy flux dissipation appears outside the photosphere, it leads naturally to prompt non-thermal emission with a large efficiency (~50%) of the total GRB energy. We will see that this high efficiency seems to be required.
- The dissipation process accelerates the flow to high Lorentz factors.
- The presence of intense magnetic fields in the dissipation region naturally explains efficient synchrotron emission.

A full reading of this Drenkhahn and Spruit review and references therein is quite illuminating for understanding the potential role of partially or completely magnetized outflows in the physics of GRBs.

So let us say a few words about the main challenger to the internal shock model; the electromagnetic model (Thompson 1994, Drenkhahn 2002, Drenkhahn & Spruit 2002, Lyutikov & Blandford 2003, Vlahakis & Königl 2003). In the introduction to their paper, Lyutikov and Blandford (2003) list 11 questions to which, in their opinion, the fireball/internal shock model does not provide satisfactory answers. So, they propose an alternative model called EMM (for Electro-Magnetic Model) illustrated in Figure 5.7; in this model the energy of a Poynting-dominated flux is dissipated directly into emitting particles through current-driven instabilities. The Poynting flux is created when a rotating, relativistic, stellar-mass progenitor loses much of its rotational energy in an electromagnetically dominated bubble which expands relativistically after breaking out from the surface of the stellar progenitor.

![Figure 5.7. Schematic illustration of the electromagnetic model (EMM) proposed by Lyutikov and Blandford (2003). See also Color section.](image-url)
A simple description of this model, with a discussion of its observable predictions has been given by Lyutikov (2006). Various aspects of Poynting-flux-dominated models have been discussed in the literature: the possible development of magnetic towers in the interior of massive rotating stars after their core has collapsed to form a black hole with an accretion disk or a millisecond magnetar (Uzdensky & MacFadyen 2006), numerical simulations of the Poynting jet (e.g. Paesold, Blackman, & Messmer 2005, McKinney. 2006, Giannios & Spruit 2006). Interestingly, the birth of Poynting-dominated jets can also be simulated in the laboratory (Lebedev et al. 2005 and references therein). One problem of the EMM is that the modeling of magnetic dissipation is very complex, making observational predictions difficult. The prompt emission offers no easy clue since the accelerated electrons radiate their energy through synchrotron emission, in the EMM as in the internal shock model (Smolsky & Usov 2000, Lyutikov & Blandford 2003, Giannios & Spruit 2006). Fortunately, the predictions of the EMM and internal shock models differ during the early afterglow phase: The EMM is characterized by a faint or absent reverse shock (Zhang & Kobayashi 2005, Lyutikov 2006) and by the existence of a bright early X-ray afterglow (see figure 3 of Genet, Daigne, & Mochkovitch 2006). Roming et al. (2006) have stressed the lack of bright optical afterglows in Swift-UVOT. They analyze the various reasons which can suppress the expected optical emission and tentatively conclude that it is due to a suppression of the strong reverse shock emission, which could be attributed to a Poynting-flux-dominated flow or to a pure non-relativistic hydrodynamic reverse shock. Genet, Daigne, and Mochkovitch (2006) studied in detail the afterglow predicted by the electromagnetic outflow models and concluded that the differences with the internal shock model are too small to permit distinguishing these two families of models from only the observation of the afterglow. Giannios (2006) argued that delayed electromagnetic dissipation naturally explains the X-ray flares discovered by Swift in the early afterglows of many GRBs. Scenarios with Poynting-flux-dominated outflows require a source of magnetic fields (>10^{15} gauss) and large rotation rates (Ω ~ 10^4 s^{-1}) (Lyutikov & Blackman 2001). These characteristics can be found when a neutron star merger forms an accretion torus surrounding a black hole and in newly born ‘magnetars’ (strongly magnetized neutron stars; Usov 1992, Blackman & Yi 1998, Thompson 2005). The strongly magnetized relativistic wind, its instability (Lyutikov & Blackman 2001) and its interaction with the ambient medium provide plausible ways of producing high-frequency emission and represent plausible mechanisms for the generation of cosmological GRBs.

Other models

- Some authors have searched for ways of attributing the entire GRB (including prompt emission) to external shocks. A way out of the angular spreading problem (which leads to smooth light-curves with no sharp spikes; Fenimore et al. 1999) is possible if the ISM is inhomogeneous and made out of small dense clouds with a large filling factor (e.g. Dermer & Mitman 1999, Dermer et al. 1999). There is some consensus, however, on the fact that these models require too many ad hoc assumptions to be realistic.
• Dar and de Rújula (2004) have developed a ‘cannonball’ model of GRBs. In the cannonball model a supernova emits bullets of matter having about the mass of the Earth. The prompt \( \gamma \)-ray emission is explained by inverse Compton scattering of the supernova light by the electrons in the cannonball’s plasma. The afterglow is successfully explained by the synchrotron radiation of ambient electrons swept up in and accelerated within the cannonball. According to the authors of the cannonball model, the only obstacle to constructing a complete theory of GRBs is the theoretical understanding of the cannonball ejection mechanism in supernova explosions. Another model involving bullets was proposed by Heinz and Begelman (1999). These authors suggested an inner engine operating like a shotgun, emitting multiple narrow bullets with an angular size much smaller than \( \Gamma^{-1} \). This shower of cold, heavy blobs with a bulk Lorentz factor \( \Gamma \sim 1000 \) runs into a dense medium which can be a strong stellar wind. The GRB is produced in the shocks that these bullets drive into the circumstellar medium. After the GRB phase the blast waves of the individual bullets merge into a single collimated shock front giving a standard afterglow in a declining external density profile. The authors show that they can reproduce the temporal structure of the prompt GRB emission with this mechanism, even for spiky bursts. This model, however, does not explain the observed temporal correlations between the duration of GRBs and their short-term variability, because these two timescales have different origins. Here again, the authors do not explain the origin of the bullets.

5.4 CONCLUSION

GRBs at cosmological distances involve a huge amount of energy liberated in a small region within a very short time, creating a fireball. We have discussed how the fireball, which is initially extremely opaque to pair creation, expands relativistically. However, such fireballs cannot explain the non-thermal spectrum of the typical GRBs and shocks have been introduced:

• Internal shocks which develop within the fast-moving ejecta. The time-varying outflow from the central engine produces successive shells ejected with different Lorentz factors.
• The external shock which appears when the outflow interacts with the external medium, at later times.

This is the framework of the popular ‘standard model’. We have defined the characteristic distances that make it possible to define the different phases of the development of the outflow. The presence in this outflow of a weak baryonic load associated with the different shocks allows transformation of the initial large bulk kinetic energy of the flow into random kinetic energy of the particles.

We have indicated that internal shocks are able to reproduce the fast temporal variability of GRB light-curves, while this is much more difficult with external shocks. Internal shocks will appear within a variable relativistic wind produced by a highly
variable source. If the Lorentz factor of the wind is variable, successive shells will have large relative velocities. Hence a fast shell injected after a slower one can eventually catch up and collide with it, giving birth to internal shocks. The electrons accelerated in these shocks emit synchrotron radiation in the presence of magnetic fields. Each shock produces its own emission and the resulting prompt emission (spectrum and variability) is the addition of simultaneous emissions from multiple shocks. The general characteristics of the prompt emission have been reproduced by different authors.

The particles accelerated in the shocks, and the presence of magnetic fields, are the ingredients needed to produce the synchrotron emission at the origin of the GRB and its afterglow. We have discussed particle acceleration and introduced the microphysical parameters which are used to characterize the ratio of the magnetic energy and of the electron energy to the energy dissipated in the shock. Then synchrotron emission has been described and the synchrotron spectrum radiated by the accelerated electrons calculated under the two conditions of fast and slow cooling. The possible contribution of inverse Compton scattering has also been discussed because it can play a significant role. These calculations are used to characterize the prompt emission. They will also play a central role in the analysis of the afterglow in Chapter 6.

But two major issues remain. The first one is the agreement between the low-energy part of GRB spectra and the optically thin synchrotron emission from shock-accelerated electrons. This problem of synchrotron emission models has been discussed with the introduction of a ‘line of death’ which is crossed by a significant fraction of GRBs. However, clever solutions have been proposed to solve the problems raised by these bursts. We have also stressed the difficulty of measuring the low-energy spectral index which is at the origin of the discrepancy. So this open issue is probably not as dramatic as is often claimed and we consider that it certainly cannot be used to question the internal shock model.

A second difficulty is perhaps more serious. It concerns the efficiency of internal shocks to convert the wind kinetic energy into hard X-ray photons. From a theoretical point of view, this efficiency cannot be very high, typically several percent. This has dramatic consequences for the energy budget needed to explain GRBs, which can reach values that are much larger than the kinetic energy of typical supernovae. Here again solutions have been proposed to increase this efficiency and we have indicated some of them. Even the question of the relevance of the measurement of the total energy released in the burst by the gamma-ray fluence has been raised. As we will see in Chapter 7, this issue will become even more crucial with the observations of early afterglows performed by Swift.

After these concerns it was natural to close this chapter with a brief outline of some models which are not based on internal shocks. One serious challenger is the electromagnetic model in which the energy of a Poynting-dominated flux is dissipated directly into emitting particles through current-driven instabilities. Another model is the cannonball model with the emission of bullets of matter whose origin is not very clear. Finally, external shocks models still have their defenders, in spite of the objections that we have reported. This provides a good transition to the next chapter,
which leaves the prompt GRB, to discuss the origin of the afterglow with its multi- 
wavelength signatures.

5.5 REFERENCES

Achterberg, A., et al. (2001) Particle acceleration by ultrarelativistic shocks: theory and 
533, L119–L122.
International Cosmic Ray Conference 11, 132–137.
Band, D., et al. (1993) BATSE observations of gamma-ray burst spectra. I—Spectral diversity, 
Bursts: 30 Years of Discovery 727, 81–85.
Bednarz, J., & Ostrowski, M. (1998) Energy spectra of cosmic rays accelerated at ultra- 
Bell, A. R. (1978a) The acceleration of cosmic rays in shock fronts. II, Monthly Notices of the 
Bell, A. R. (1978b) The acceleration of cosmic rays in shock fronts. I, Monthly Notices of the 
Beloborodov, A. M. (2000) On the efficiency of internal shocks in gamma-ray bursts, 
Berger, E., et al. (2003) A common origin for cosmic explosions inferred from calorimetry of 
Fluids 19, 1130–1138.
Blandford, R. D., & Ostriker, J. P. (1978) Particle acceleration by astrophysical shocks, 
blackbody spectrum look non-thermal?, Astronomy Reports 43, 739–747.
Cavallo, G., & Rees, M. J. (1978) A qualitative study of cosmic fireballs and gamma-ray bursts, 
Cohen, E., et al. (1997) Possible evidence for relativistic shocks in gamma-ray bursts, 
Crider, A., et al. (1997) Evolution of the low-energy photon spectral in gamma-ray bursts, 


6

Gamma-ray burst afterglows

6.1 THE DYNAMICS OF THE FIREBALL

6.1.1 Evolution of the fireball

Initially, the fireball is extremely hot (its pressure $p$ is much greater than $\rho$, the rest mass density of the fluid element). The Lorentz factor $\Gamma$ of the narrow shell with a radial width $\Delta$ increases linearly with its radius: $\Gamma \propto R$. At a radius $R_{\text{sat}} = \eta R_{\text{in}}$ the shell is cold and it coasts at constant $\Gamma = \Gamma_{\text{max}} = \eta$ ($\eta$ is the energy-to-mass ratio in the initial fireball or the mean energy per baryon: $\eta = E_0 / M_0 c^2$ where $M_0 \ll E_0 / c^2$ represents the baryon load of the fireball). The internal energy of the fireball has been entirely converted into kinetic energy of the baryons. This is the end of the acceleration phase. The internal energy of the fireball is now negligible compared to the rest mass energy ($p \ll \rho$). The fireball behaves like a pulse of energy, with a frozen radial profile, which has almost the speed of light. We have seen that this pulse approximation breaks down at $R_{\text{s}} \sim R_{\text{in}} \eta^2$. At this distance, the fireball begins to spread. Since it is inhomogeneous with a non-monotonic distribution of velocities as a function of radius, internal shocks can be produced. These internal shocks convert part of the kinetic energy of the fireball into radiation (see Figure 5.1). At the beginning of the fireball expansion the interstellar medium has no influence on the expanding shell. But as the cold shell propagates with a Lorentz factor $\Gamma$ it drives a shock into the interstellar medium (ISM). This shock propagates with $\Gamma_{\text{sh}} = 2^{1/2} \Gamma$. Behind the shock the ISM is heated. As the shell radius $R$ increases, more ISM matter is shocked and the shell is progressively influenced by the ISM. This influence becomes significant when the energy of the heated ISM becomes comparable to $E_0$. The afterglow emission begins when most of the bulk kinetic energy of the ejecta is transferred to the shocked external medium. For long GRBs this can take place while the GRB is still going on. Figure 6.1 gives an artist’s view of this model with its
Figure 6.1. An artist’s view illustrating the various steps of the basic standard model with the internal and external shocks and the various radiations they emit (from Gehrels, Piro, & Leonard 2007). On the left are indicated the two main scenarios which lead to a central black hole surrounded by a disk. These two types of progenitors are discussed in Chapter 8. See also Color section.

Two main phases: the internal shocks and the external shock with the ambient medium. Figure 6.2 illustrates the regions involved in the basic standard model.

The rest-frame thermal energy of the material swept up by the blast wave is $\Gamma mc^2$, where $m$ is the mass of the shocked material. In our frame this energy is blueshifted so that it is $\Gamma^2 mc^2$. As the radius of the fireball increases an increasing mass is swept up ($m \propto R^3$). The kinetic energy loss of the shell becomes significant and the shell deceleration starts ($m_{\text{dec}} = (4\pi/3)\rho_{\text{ext}}R_{\text{dec}}^3$). Precisely, the deceleration radius $R_{\text{dec}}$ has been defined as the radius at which the fastest (initial) part of the ejecta moving with Lorentz factor $\Gamma_{\text{max}} = \eta$ sweeps up an amount of external gas equal to a fraction $\Gamma_{\text{max}}^{-1}$ of its own mass: $M_0 = E_0/\Gamma_{\text{max}}c^2 = E_0/\eta c^2$ (Panaitescu & Mészáros 1999). Hence $m_{\text{dec}} = M_0/\eta = E_0/\eta^2 c^2$. This is of course much sooner than in the non-relativistic case where $m_{\text{dec}} = M_0$. Beyond this radius the shocked gas dominates
Figure 6.2. Schematic view of the phases involved in the fireball model with internal and external shocks. The left-hand scale gives the logarithm of the distance from the central object producing the GRB (meters). The right-hand scale indicates the various phases: (1) Phase of acceleration: the thermal or magnetic energy is converted into kinetic energy of the baryons within the ejecta. The ejecta emitted in a jet characterized by its beaming angle becomes ultrarelativistic. (2) End of the acceleration phase, the Lorentz factor of the ejecta reaches its maximum value, typically a few hundred. (3) The ejecta become transparent. We have already defined the photospheric radius; if internal energy is available it can be radiated as quasi-thermal emission. (4) Internal shocks (CI) appear and progress within the relativistic ejecta. The matter heated by the shocks is the source of the GRB. (5), (6), (7) These phases characterize the various parts of the external shocks, which start when the deceleration due to the circum-burst medium appears: (5) corresponds to the reverse shock (CR) which propagates within the ejecta; (6) is the contact discontinuity (see also Figure 6.3 for the different regions of the external shock); (7) is the forward shock (CA) that is responsible for the afterglow with its emission from radio to gamma-ray wavelengths. (8) Corresponds to a region where the Lorentz factor has greatly decreased and the expansion becomes non-relativistic. Depending on the kinetic energy of the ejecta and the density of the circum-burst medium, the two main phases of internal shocks and external shocks can be well separated or quasi-simultaneous (Atteia & Mochkovitch 2004). See also Color section.
Figure 6.3. Upper panel: Schematic representation of the regions that can be identified when a relativistic fireball interacts with the interstellar medium (ISM). The solid line indicates the density as a function of radius. Zone 1 is the undisturbed ISM at large distances; Zone 4 is the unshocked cold shell at a small distance from the central engine. In between these two regions, zones 2 and 3 are the shocked regions separated by the contact discontinuity (CD). On each side the pressure and the Lorentz factor are identical but not the density. Zone 2 is the region of the ISM that has been heated by the forward shock. Zone 3 is the part of the shell heated by the reverse shock (from Sari, Narayan, & Piran 1996). Lower panel: The Lorentz factor ($\gamma$), density ($\rho$), and pressure ($p$) in the four regions derived from simulations performed by Kobayashi, Piran, and Sari (1999).
the mass and energy of the expanding system. In the early adiabatic stage of fireball expansion, the thermal energy of the fireball is converted into bulk kinetic energy of expansion ($\Gamma$ increases with time), and when deceleration becomes significant the process is reversed and the fireball starts reconverts its bulk kinetic energy into thermal energy. The fireball kinetic energy is used to heat the swept up external medium (see Figure 6.3). This thermal energy is released in an optically thin region and it can be radiated away if cooling times are short enough (Rees & Mészáros 1992).

If radiative losses are unimportant (adiabatic evolution), $\Gamma$ slowly starts to decrease as $\Gamma \propto m^{-1/2}$, since $E \propto \Gamma^2 m = cst$. In the case of a uniform ambient medium with $\rho = cst$ and $m \propto R^3$ we have $\Gamma \propto R^{-3/2}$. If radiative cooling is important, it carries away all the re-randomized energy, and $\Gamma \propto R^{-3}$ (see Section 6.1.2).

The emission received by an observer from a spherical shell at radius $R$ moving with a bulk Lorentz factor $\Gamma$ is restricted to the part of the shell that is moving close to the line-of-sight, with an angle smaller than $\theta = \Gamma^{-1}$ (see Figure 5.3). An impulsive energy release at the shell is thus observed over an angular time scale $t \sim R/2c\Gamma^2$ (see, for example, Rees & Mészáros (1992) for this association between the observer time $t$ and the distance $R$, and Section 5.2.2). Assuming an impulsive ejection, the bulk of the ejected matter at a radius $R$ is mainly concentrated in a region of width $\Delta R \sim R/2\Gamma^2$. The timescale over which the deceleration is observed to take place is given by $t_{\text{dec}} \sim R_{\text{dec}}/2c\Gamma^2$, where $R_{\text{dec}}$ is the deceleration radius. This is the observer timescale over which the external shock radiation will be detected for an impulsive ejection. In the next section we derive the dependence of the bulk Lorentz factor $\Gamma$ versus $R$, and the time-dependence of $R$. In the adiabatic and radiative regimes one respectively finds $R \propto t^{1/4}$ and $R \propto t^{1/7}$, leading to $\Gamma \propto R^{-3/2} \propto t^{-3/8}$ and $\Gamma \propto R^{-3} \propto t^{-3/7}$. These are the time dependence of $R$ and $\Gamma$ after the characteristic deceleration radius $R_{\text{dec}}$. As can be seen, the radius is changing much more slowly with time than the Lorentz factor. So the system can be viewed as almost standing still with the Lorentz factor decreasing with time.

### 6.1.2 The temporal evolution of $R$ and $\Gamma$

Various authors have computed the expected evolution of $R$ and $\Gamma$ with time. We follow below the discussion of Piran (1999), which is valid when the relativistic jet interacts with a constant density (homogeneous) medium. The homogeneous shell is described by its rest-frame energy $M$ (rest mass and thermal energy) and its Lorentz factor $\Gamma$. Initially, $E_0 = M_0c^2$ and $\Gamma_{\text{max}} = M_0c^2 \eta$. The shell collides with the surrounding medium (ISM). The mass of the ISM that has already collided with the shell is $m(R)$. As the shell propagates it sweeps up more ISM mass characterized by $dm$.

The additional mass element $dm$, which was at rest, collides inelastically with the shell. Energy and momentum conservation yield:

$$\frac{d\Gamma}{\Gamma^2} - 1 = -\frac{dm}{M}$$
Figure 6.4. Fast-cooling synchrotron spectra from a shock-accelerated population of electrons. The ejecta are assumed to follow a spherical adiabatic evolution and the electrons to have a power-law distribution of energies of index $p$. Fast cooling is expected at early times (prompt GRB and early afterglow). The shape of the spectrum is determined by the ordering of the self-absorption frequency, $\nu_{sa}$, with respect to $\nu_c < \nu_m$. There are three possible shapes for the spectrum, corresponding to $\nu_{sa} < \nu_c$, $\nu_c < \nu_{sa} < \nu_m$, and $\nu_{sa} > \nu_m$, from top to bottom. The middle panel ($\nu_c < \nu_{sa} < \nu_m$) illustrates the “canonical” situation, which arises for a reasonable choice of parameters for afterglows in an ISM environment. Scalings in both ISM and wind environments are given only in the top two panels since, for typical parameters, the situation of the bottom panel ($\nu_{sa} > \nu_m$) does not happen during the FC phase of the afterglow (from Granot, Piran, & Sari 2000).

where $Mc^2$ is the rest-frame energy of the shell (rest mass + thermal energy)

$$dE/c^2 = (\Gamma - 1) \, dm$$

and where $dE$ is the thermal energy produced in the collision. If $\varepsilon$ is the fraction of the shock-generated thermal energy (relative to the observer frame) that is radiated, it
results that
\[ dM = (1 - \varepsilon) \frac{dE}{c^2} + dm = [(1 - \varepsilon) \Gamma + \varepsilon] \, dm \]

Piran (1999) deduces relations between the Lorentz factor and the total mass of the shell and between \( m(R) \) (and therefore \( R \)) and \( \Gamma \) describing the hydrodynamical evolution of the shell.

In the limit where \( \Gamma_{\text{max}} \gg \Gamma \gg 1 \) he obtains various results:

1. First by solving the analytic relation between \( m(R) \) and \( \Gamma \), a well-known result is obtained: the Lorentz factor decreases and reaches \( \Gamma = \Gamma_{\text{max}}/2 \) when the swept-up mass is \( m \sim M_0/2\Gamma_{\text{max}} \).
2. The second result concerns the expression of \( m(R) \):

\[ m(R) = \left[ M_0/(2 - \varepsilon) \Gamma_{\text{max}} \right] (\Gamma/\Gamma_{\text{max}})^{(2+\varepsilon)} \]

As \( m(R) \propto R^3 \) and \( M_0 \), \( \Gamma_{\text{max}} \), and \( \varepsilon \) are fixed, it results that \( R^3 \propto \Gamma^{-2+\varepsilon} \) or \( \Gamma \propto R^{-3/(2-\varepsilon)} \).

With this simple relation the two important regimes, adiabatic and radiative, can be studied to derive the evolution of \( R \) and \( \Gamma \) with time.

In the adiabatic case, the energy \( E \) of the shock is approximately constant. The energy carried away by the radiated photons must be a negligible fraction of the total energy of the ejecta. This assumption is correct if the energy density of the electrons accelerated at the shock front is a small fraction \( \varepsilon_e \ll 1 \) of the total energy density in the post-shock fluid, or if most of the electrons are adiabatic, i.e. their radiative cooling timescale exceeds that of the adiabatic losses due to the remnant expansion (slow cooling).

So, in the adiabatic case \( \varepsilon \sim 0 \) and \( \Gamma \propto R^{-3/2} \) (Blandford & McKee 1976, Katz 1994b, Wijers, Rees, & Mészáros 1997, Waxman 1997a, b, c). As we have seen, the observer time corresponding to a given radius and Lorentz factor is

\[ t_{\text{obs}} = \frac{R}{2\Gamma^2 c} \propto R^{(1+3)} \propto R^4 \]

The dependence of \( R \) and \( \Gamma \) can thus be given versus time: \( R(t) \propto t^{1/4} \) and \( \Gamma(t) \propto t^{-3/8} \).

In the fully radiative case, the radiation losses are fast compared to the hydrodynamic timescale and all the internal energy in the shock is radiated away. This requires two conditions: the fraction of the energy which goes into the electrons has to be large (\( \varepsilon_e \to 1 \)), and the regime must be fast cooling.

For the fully radiative regime with \( \varepsilon = 1 \) we have \( \Gamma \sim R^{-3} \) and using the expression for \( t_{\text{obs}} \) we obtain: \( t_{\text{obs}} \propto R^{(1+6)} \propto R^7 \) leading to \( R(t) \propto t^{1/7} \) and \( \Gamma(t) \propto t^{-3/7} \) (Blandford & McKee 1976, Vietri 1997, Katz & Piran 1997). The conditions needed for this regime seem rather improbable and it is not expected to observe many afterglows in this regime.

In order to completely define the two functions \( R(t) \) and \( \Gamma(t) \), Sari (1997) has rewritten the self-similar solution that describes the adiabatic slowing down of an extremely relativistic shell propagating into the ISM established by Blandford and
McKee in 1976. He gives the expressions of \( n(r, t), \Gamma(r, t), e(r, t) \) which are respectively the density, the Lorentz factor and the energy density of the material behind the shock. \( n(r, t) \) and \( e(r, t) \) are measured in the fluid’s rest frame while \( \Gamma(r, t) \) is relative to an observer at rest.

The scaling laws giving \( R(t) \) and \( \Gamma(t) \) can be expressed as a function of \( E = E_0 \), the total energy in the adiabatic flow and \( n \) the density of the interstellar medium. He obtains (Sari 1997):

\[
R(t) = 3.2 \times 10^{16} E_{52}^{1/4} n_1^{-1/4} t_s^{1/4} \text{ cm}
\]

\[
\Gamma(t) = 260 E_{52}^{1/8} n_1^{-1/8} t_s^{-3/8}
\]

where \( E_{52} \) is the energy measured in units of \( 10^{52} \) erg, \( n_1 \) is the ISM density in units of \( 1 \text{ cm}^{-3} \), and \( t_s \) is the time in seconds. Such a solution can be used as the starting point for detailed calculations of the emitted radiation.

As the Blandford and McKee solution is for the adiabatic regime, it does not allow taking energy losses into account. Cohen, Piran, and Sari (1998) have derived a self-similar radiative solution in which an arbitrary fraction of the energy generated in the shock is radiated away. The radiation process is assumed to be fast, taking place only in the vicinity of the shock and radiating away a fixed fraction of the energy generated by the shock. Their fully radiative extreme relativistic solution is different from the Blandford and McKee fully adiabatic relativistic solution. Such a solution can be used to describe the evolution of the GRB afterglow in its non-adiabatic stage. In the early stages of the afterglow, the electron cooling is fast since the density behind the shock is low, the coupling between the electrons and the protons is only partial, and the protons’ energy is not radiated away. This is why the shock is semi-radiative rather than fully radiative. This is exactly the scenario described in the paper of Cohen, Piran, and Sari (1998).

The characteristics of the synchrotron emission, its spectrum and light-curve, can be calculated using \( R(t) \) and \( \Gamma(t) \). This will be done in Sections 6.3 and 6.4.

6.2 EXTERNAL SHOCKS

The external shocks appear when a relativistic shell is slowed down in the cold interstellar medium (ISM) or in the stratified wind ejected by the progenitor star prior to the explosion (Rees & Mészáros 1992). Generally two shocks form: an outgoing shock (usually called the external shock) that propagates into the surrounding medium, and a reverse shock that propagates into the ejecta. The afterglow begins at a distance where most of the energy of the ejecta is transferred to the medium. For a long burst this can take place while the GRB is still going on.

At the very beginning as the external shock builds up, its bolometric luminosity \( L \) rises approximately as \( L \propto t^2 \), if the surrounding medium is the ISM with a constant density. This is obtained equating, in the contact discontinuity frame, the kinetic flux \( L/4\pi R^2 \) to the external pressure \( p_{\text{ext}} \Gamma^2 \), during the initial phase while \( \Gamma = \Gamma_f = c/\sqrt{\gamma} \), \( p_{\text{ext}} \) being also const. \( R \propto t \) and \( L \propto R^2 \) give \( L \propto t^2 \). The luminosity increases for as
long as the external material can be neglected. It peaks when $R$ reaches the typical
deceleration radius, and decays rapidly thereafter (Rees & Mészáros 1992, Sari 1997).
We will see below that the dynamical evolution of the afterglow and its radiative
properties are fully determined by $E_0$, $\varepsilon_e$, $\varepsilon_B$, $n$, and $p$, the spectral index of the shock-
accelerated electrons (see for instance Panaitescu & Kumar 2000). In particular it is
independent of $\eta$ which characterizes the baryon load of the fireball.

The forward and reverse shocks delimit four regions illustrated in Figure 6.3:
region 1 is the ISM, region 2 is the shocked ISM between the contact discontinuity
(CD) and the forward shock (FS), and region 3 is the shocked shell between the
contact discontinuity and the reverse shock (RS). The reverse shock is situated in
contact with the unshocked shell, which is region 4. Pressures and velocities stay
equal on each side of the contact discontinuity, but this is not the case for the density
$\rho$. Kobayashi, Piran, and Sari (1999) give the values of the Lorentz factor $\Gamma$, of the
density $\rho$, and of the pressure $p$ on both sides of the contact discontinuity.

The emitted radiation and the cooling efficiency depend on the conditions in the
shocked regions 2 and 3. The shock structure is determined by two quantities:

- $\Gamma$, the bulk Lorentz factor of the inner expanding matter (region 4) relative to the
  outer matter (the ISM, region 1);
- the ratio $n_4/n_1$ between the particle number densities in these two regions.

Initially, during the early phase of the external shock, the ratio $n_4/n_1$ is large
$n_4/n_1 > \Gamma^2$, and the expanding shell is small and dense. During this early phase
the reverse shock is mildly relativistic and most of the energy conversion takes place
in the forward shock. Only a very small fraction of the energy is converted into
thermal energy in the reverse shock (Sari & Piran 1995). The values of $\Gamma$, $n$, and $e$, the
particle and energy densities in the shocked regions have been given by Piran (1999)
for instance.

The Sedov length is often introduced. It is defined by:

$$E_0 = m_p c^2 \int_0^l 4\pi r^2 n(r) \, dr$$

The rest mass energy within the Sedov sphere equals the energy of the explosion $E_0$.

For a homogeneous ISM, we have the following equations:

$$E_0 = m_p c^2 n \int_0^l 4\pi r^2 \, dr = m_p c^2 n \left( \frac{4\pi}{3} \right) l^3$$

and

$$l = \left[ \frac{3E_0}{4\pi m_p c^2 n} \right]^{1/3} = 10^{18} E_0^{1/3} n_1^{-1/3} \text{ cm}$$

A second length scale must be introduced, $\Delta$, the initially narrow width of the
relativistic shell $\Delta \sim R_{\text{in}}$ with values in the range $\Delta \sim 10^7$ to $10^{12}$ cm (Sari &
Piran 1995, 1997) in the observer’s rest-frame. With these two characteristic
lengths Sari and Piran (1995) define the dimensionless parameter,
\[ \xi = (l/\Delta)^{1/2} \Gamma^{-4/3} = 2(l_{18}/\Delta_{12})^{1/2} \Gamma^{-4/3} \]

The canonical values for \( \Delta \) and \( l \) are 10\(^{12} \) cm and 10\(^{18} \) cm, respectively. Two radii, \( R_{\text{dec}} \) and \( R_\Delta \), are defined to represent the locations where the effective energy extraction takes place.

Half of the shell’s kinetic energy is converted into thermal energy when the collected external mass is \( M_0/\Gamma \). This takes place at a characteristic distance \( R_{\text{dec}} \), which has been called the deceleration radius

\[ R_{\text{dec}} \propto l/\Gamma^{2/3} = \left( \frac{3E_0}{4\pi n_m c^2 \Gamma^2} \right)^{1/3} = 5.4 \times 10^{16} E_{52}^{1/3} n_1^{-1/3} \Gamma_{100}^{-2/3} \text{ cm} \]

where \( E_{52} \) is the equivalent isotropic energy in units of 10\(^{52} \) erg, and \( n_1 \) the density in particles per cubic centimeter.

The other characteristic distances are \( R_\Delta \) and \( R_N \):

- \( R_\Delta \) is the distance at which the reverse shock reaches the inner edge of the shell (Sari & Piran 1995): \( R_\Delta = l^{3/4} \Delta^{1/4} = 10^{15} l_{18}^{3/4} \Delta_{12}^{1/4} \) cm.
- \( R_N \) is the distance at which the energy density produced by the shock becomes high enough for the reverse shock to become relativistic and to start reducing the Lorentz factor of the shell considerably: \( R_N = l^{3/2} \Delta^{1/2} \Gamma^2 \).

A simple relation can be obtained between these characteristic radii using \( \xi \) (Kobayashi, Piran, & Sari, 1999):

\[ R_\Delta \sqrt{\xi} = R_{\text{dec}} = R_N/\xi \]

External shocks become effective at \( \min(R_{\text{dec}}, R_\Delta) \) and typical values for \( l, R_{\text{dec}}, \) and \( R_\Delta \) are: \( l \sim 10^{18} \) cm, and \( R_{\text{dec}} \sim R_\Delta \sim 10^{16} \sim 10^{17} \) cm. Conditions on \( R_\Delta, R_{\text{dec}}, \) and on the value of the parameter \( \xi \) are used by Sari and Piran (1995) to define when the reverse shock is Newtonian (\( \xi > 1 \)) or relativistic (\( \xi < 1 \)). Both cases are possible, depending on \( \Delta \) and \( \Gamma \), which are the two quantities which are less constrained, whereas the Sedov length is well defined, \( l = 10^{18} \) cm. The case \( \xi = 1 \) corresponds to typical values of \( l, \Delta, \Gamma \) (Piran 2005), leading to \( R_\Delta = R_{\text{dec}} = R_N \).

Piran (2005) has defined the conditions at the forward and reverse shocks and the magnetic field in regions 2 and 3, when the shell has expanded to \( R \). The shock heats the expanding ejecta, amplifying pre-existing magnetic fields or generating a turbulent field and accelerating the electrons. The energy distribution of the relativistic electrons and the magnetic field behind the shock are needed to calculate the synchrotron spectrum. To characterize the energy carried by the electrons and by the magnetic field, two numbers are used: \( e_e \) and \( e_B \). It is commonly assumed that \( e_e \) and \( e_B \) represent a constant fraction of the internal energy behind the shock (called \( e \)). It is thus possible to define two dimensionless parameters:

\[ \varepsilon_e = e_e/e \quad \text{and} \quad \varepsilon_B = B^2/8\pi = e_B/e \]

\( \varepsilon_e \) and \( \varepsilon_B \) are generally supposed to stay constant throughout the burst evolution. This is a simplification because the evolution of these parameters during the fireball
expansion is certainly complex and difficult to calculate. The problem of the magnetic field strength and its evolution, in particular, is a serious concern (Thompson 1994, Mészáros, Laguna, & Rees 1993).

For the typical case with \( \xi = 1 \), we can compute the physical parameters of the shocked regions at \( R_{\Delta} = R_{\text{dec}} \), when the bulk kinetic energy of the shell is converted to thermal energy via the two shocks. We have for the forward shock, \( \Gamma_2 = \Gamma \), \( n_2 = 4\Gamma n_1 \), and \( e_2 = 4\Gamma^2 n_1 m_p c^2 \), and, for the reverse shock, \( \Gamma_3 = \Gamma \), \( n_3 = 4\Gamma^2 n_1 \), and \( e_3 = e_2 \) at the contact discontinuity; and, for the magnetic field, \( B \propto \varepsilon_B^{1/2} \Gamma n_1^{1/2} \) (Sari, Narayan, & Piran 1996). These authors also give more complete relations including the \( \xi \)-dependence.

Having computed the physical parameters in the shocked regions it is now possible to calculate their synchrotron radiation. When they studied the evolution of relativistic cosmological fireballs, many authors predicted that they would be followed by afterglows radiating at frequencies gradually declining from soft \( \gamma \)-rays through X-rays to visible light and radio waves. Moreover, these afterglows would have to last for a time significantly longer than the burst itself (Katz 1994a,b, Mészáros & Rees 1997, Vietri 1997, Sari 1997, Waxman 1997b, Wijers, Rees, & Mészáros 1997). The equations given in this section permit the calculation of the evolution of the Lorentz factor \( \Gamma \) with the observer time \( t \). It is necessary to know this evolution for the calculation of the synchrotron emission of the electrons accelerated by the forward shock, as we explain in the next section.

### 6.3 ENERGY SPECTRUM OF THE AFTERGLOW

The bulk of the kinetic energy of the shell is converted into thermal energy via the two shocks at around the time the shell has expanded to the radius \( R_{\Delta} \) defined in the previous section.

Using \( R(t) \) and \( \Gamma(t) \), the shock conditions which connect density and energy in the shocked regions \( (n_2 \sim 4\Gamma n_1, \; e_3 = e_2 = e = 4\Gamma^2 n_1 m_p c^2 \) and \( n_1 = 7n_4 \) and of course the values of \( \varepsilon_e \) and \( \varepsilon_B \), the synchrotron spectrum and the characteristic frequencies can be calculated. Sari, Piran, and Narayan (1998), for instance, give the synchrotron spectra expected for fast and slow cooling (Figure 5.4). The synchrotron spectrum can be represented by segments of power laws between the characteristic frequencies (Figures 5.4 and 6.4). The following simple analysis can be carried out to explain the frequency-dependence of \( F_\nu \), starting from the lowest frequencies of the spectrum.

In the Rayleigh–Jeans part of the spectrum that is optically thick to synchrotron self-absorption we have a blackbody-like radiation, with \( F_\nu \propto kT_{\text{eff}} \nu^2 \).

For slow cooling \( kT_{\text{eff}} \sim \gamma m_e c^2 = c\gamma t \) throughout the whole shell of shocked fluid behind the shock. Therefore \( F_\nu \propto \nu^2 \) below \( \nu_{\text{sa}} \), the self-absorption frequency.

For fast cooling in the optically thick regime \( (\nu < \nu_{\text{sa}}) \), the Rayleigh–Jeans portion of the black-body spectrum is defined by \( F_\nu \propto \nu^2 \gamma_{\text{typ}}(\nu) \), where \( \gamma_{\text{typ}}(\nu) \) is the typical Lorentz factor of the electrons (Granot, Piran, & Sari, 2000). As most of the photons are emitted at optical depth \( \tau_\nu \sim 1 \), \( \gamma_{\text{typ}}(\nu) \) has to be evaluated at the place
where $\tau_\nu \sim 1$. As $\nu$ decreases below $\nu_{sa}$ the location where $\tau_\nu \sim 1$ moves from the back of the shell toward the front where the electrons suffer less cooling, so that $kT_{\text{eff}}(\tau_\nu = 1) \propto \nu^{-5/8}$ (Granot, Piran, & Sari 2000, Piran 2005) and $F_\nu \propto \nu^{11/8}$. This corresponds to the fast-cooling spectrum of an inhomogeneous electron temperature distribution. The black-body spectrum is modified by the fact that the effective temperature varies with frequency.

In an ongoing shock there is a continuous supply of newly accelerated electrons. These electrons are injected right behind the shock, with Lorentz factors $\gamma > \gamma_m$ (the spectrum of accelerated electrons is a power law $N(\gamma) \propto \gamma^{-p}$ for $\gamma > \gamma_m$) and then they begin to cool because of radiation losses. Just behind the shock there is thus a thin layer where the electrons are hot because they have not had sufficient time to cool significantly. This hot layer is very thin. Behind this thin layer there is a much wider layer of electrons that have cooled. At a certain frequency $\nu_{ac}$, $\tau_\nu = 1$ at the location behind the shock where electrons with $\gamma_m$ start to cool significantly. Below this frequency the effective temperature is constant at the place where $\tau_\nu = 1$, and we have $F_\nu \propto \nu^{-2}$ for $\nu < \nu_{ac}$ and $F_\nu \propto \nu^{11/8}$ for $\nu_{ac} < \nu < \nu_{sa}$ (see, for instance, Piran (2005) for the expression of $\nu_{sa}$ for slow cooling and $\nu_{ac}$ and $\nu_{sa}$ for fast cooling, and Figure 6.4).

Above this optically thick part of the synchrotron spectrum, the lowest part of the optically thin spectrum is always the sum of the contributions of the tails of all the electron emissions $F_\nu \propto \nu^{1/3}$. This is typical of synchrotron radiation, and it is independent of the presence of a high-energy tail in the particle distribution function (Mészáros & Rees 1993, Katz 1994b, Cohen et al. 1997). This is also independent of the exact shape of the electron distribution.

At the high-frequency end of the synchrotron spectrum $F_\nu \propto \nu^{-p/2}$. This can be explained by the fact that this part corresponds to the rapid cooling of the most energetic electrons. These electrons emit practically all their energy $\gamma m_e c^2$ at their synchrotron frequency. Since the distribution of electrons is $N(\gamma) d\gamma \sim \gamma^{-p} d\gamma$, the number of electrons with Lorentz factor $\sim \gamma$ is $\propto \gamma^{1-p}$. The energy of these energetic electrons is $\sim \gamma^{2-p}$. As these electrons cool they deposit most of their energy into a frequency range $\nu_{\text{syn}}(\gamma) \propto \gamma^2$ and $F_\nu \propto (\gamma^{2-p} / \gamma^2) = \gamma^{-p}$, and $F_\nu \propto \nu^{-p/2}$.

In the intermediate frequency region the spectrum depends on the cooling regime. For slow cooling the typical electrons with $\gamma_{\text{min}}$ do not cool on the hydrodynamic timescale (it is the opposite for fast cooling).

The critical parameter to determine the electron regime is $\gamma_c$, the Lorentz factor of an electron that cools on a hydrodynamic timescale. Two frequencies are introduced, which have a key role:

$$\nu_m = \nu_{\text{syn}}(\gamma_{\text{min}})$$

$$\nu_c = \nu_{\text{syn}}(\gamma_c)$$

*For slow cooling*, $\nu_c > \nu_m$, and $F_\nu \sim \nu^{(1-p)/2}$ for the intermediate region between $\nu_m$ and $\nu_c$ (Waxman 1997b).

*For fast cooling*, $\nu_c < \nu_m$, and $F_\nu \sim \nu^{-1/2}$ between these two frequencies. All the electrons in the system contribute as they cool faster than the dynamical timescale. As the energy of the electron is proportional to $\gamma$ and the typical frequency of
synchrotron emission is proportional to $\gamma^2$, the flux per unit frequency is proportional to $\gamma^{-1}$ or $\nu^{-1/2}$.

To summarize, if we consider the synchrotron spectrum $F_\nu \propto \nu^{-\beta}$, its various components are characterized by the following values of $\beta$ (Sari, Piran, & Narayan 1998): $-2$ for $\nu < \nu_{sa}$ self-absorption (see above the two possibilities when $\nu < \nu_{sa}$); $-1/3$ for $\nu_{sa} < \nu < \min(\nu_m, \nu_c)$; $1/2$ for $\nu_c < \nu < \nu_m$ fast cooling; $(p - 1)/2$ for $\nu_m < \nu < \nu_c$ slow cooling; $p/2$ for $\max(\nu_m, \nu_c) < \nu$.

In this combination of power laws three are fixed and two are related to the power-law index of the electron spectrum. The behavior of $F_\nu$ for fast and slow cooling is thus more complex than the simple case illustrated in Figures 5.4 and 6.4, and it depends on the ordering of the characteristic frequencies. A comparison with the spectrum of the well-known GRB 970508 measured 12.1 days after the burst has been given in Figure 3.7. The agreement is remarkable (Galama et al. 1998).

Granot and Sari (2002) provided a more complete description of all possible spectra for GRB afterglows with the identification of five possible regimes depending on the ordering of the break frequencies: two regimes for fast cooling and two for slow cooling. Another regime is identified for both fast and slow cooling when $\nu_{sa} > \nu_m, \nu_c$. Another type of spectrum for GRB afterglows in the cases of slow and fast cooling is given by Sari and Esin (2001), taking into account the presence of strong inverse Compton cooling (their Figure 1). The differences in the time evolution of cooling breaks and the slopes of the different segments of the spectrum are analyzed. They show how the physical interpretation of the observed characteristics of the synchrotron spectrum must be modified when inverse Compton cooling is at work.

### 6.4 AFTERGLOW LIGHT-CURVES IN THE STANDARD MODEL

We have discussed the instantaneous synchrotron spectrum expected in various situations. This instantaneous spectrum does not depend on the hydrodynamic evolution of the shock. The light-curves at a given frequency, however, depend on the temporal evolution of various quantities such as the break frequencies $\nu_m$ and $\nu_c$ and the peak flux $F_{\nu,\text{max}}$. These in turn depend on how $\Gamma$, the Lorentz factor of the shocked fluid, and $N_e$, the total number of swept-up electrons in the post-shock fluid, scale as a function of $t$ (Mészáros & Rees 1997, Sari, Piran, & Narayan 1998). Sari, Piran, and Narayan (1998) computed the time-dependence of $\nu_c, \nu_m, F_{\nu,\text{max}}$ for adiabatic and fully radiative evolutions, and a constant density ISM. For an adiabatic evolution they found $\nu_c \propto t^{-1/2}$, $\nu_m \propto t^{-3/2}$, and $F_{\nu,\text{max}}$ independent of time, with:

$$\nu_c = 2.7 \times 10^{12} \varepsilon_B^{-3/2} E_{52}^{-1/2} n_1^{-1/2} t_d^{-1/2} \text{ Hz}$$

$$\nu_m = 5.7 \times 10^{14} \varepsilon_B \varepsilon_e^{-1} E_{52}^{-1/2} D_{28}^{-3/2} \text{ Hz}$$

$$F_{\nu,\text{max}} = 1.1 \times 10^5 \varepsilon_B^{1/2} E_{52} n_1^{1/2} D_{28}^{-2} \mu\text{Jy}$$
where \( t_d \) is the time in days, \( n \) the density in units of \( \text{cm}^{-3} \), and \( E_{52} = E/10^{52} \text{erg} \), and \( D_{28} = D/10^{28} \text{cm} \).

For an adiabatic hydrodynamic evolution, we thus have the well-known result that the flux at the peak of the synchrotron spectrum is independent of time, in the slow-cooling (Katz 1994b, Mészáros & Rees 1997) and fast-cooling regimes (Sari, Piran, & Narayan 1998).

For the fully radiative case these evolutions are \( \nu_c \propto t^{-2/7} \), \( \nu_m \propto t^{-12/7} \), and \( F_{\nu,\text{max}} \propto t^{-3/7} \), with:

\[
\begin{align*}
\nu_c &= 1.3 \times 10^{13} \varepsilon_B^{-3/2} E_{52}^{-4/7} \Gamma_2^{-4/7} n_1^{-13/14} t_d^{-2/7} \text{ Hz} \\
\nu_m &= 1.2 \times 10^{14} \varepsilon_B^{-1/2} \varepsilon_\epsilon^{-2} E_{52}^{-4/7} \Gamma_2^{-4/7} n_1^{-1/14} t_d^{-12/7} \text{ Hz} \\
F_{\nu,\text{max}} &= 4.5 \times 10^{3} \varepsilon_B^{1/2} E_{52}^{8/7} \Gamma_2^{-8/7} n_1^{-5/14} D_{28}^{-2} t_d^{-3/7} \mu\text{Jy}
\end{align*}
\]

where \( \Gamma \), which we have also called \( \Gamma_{\text{max}} \), is the initial Lorentz factor of the ejecta and \( \Gamma_2 = \Gamma_{\text{max}}/100 \).

The cooling frequency \( \nu_c \) decreases with time slower than \( \nu_m \): \( \nu_c \propto t^{-1/2} \) and \( \nu_m \propto t^{-3/2} \) for the adiabatic regime, and \( \nu_c \propto t^{-2/7} \) and \( \nu_m \propto t^{-12/7} \) for the radiative regime. Since at sufficiently early times \( \nu_c < \nu_m \) (fast cooling) while at later times \( \nu_c > \nu_m \) (slow cooling), there is a transition between the two regimes when \( \nu_c = \nu_m \) at a time \( t_0 \). The critical frequency is \( \nu_0 = \nu_c(t_0) = \nu_m(t_0) \) (see Sari, Piran, & Narayan (1998), for the expression of \( t_0 \) in the adiabatic and radiative cases). At \( t = t_0 \) the spectrum changes from fast cooling to slow cooling. In addition if \( \varepsilon_\epsilon \rightarrow 1 \) the hydrodynamic evolution changes at this stage from radiative to adiabatic (Mészáros, Rees, & Wijers 1998). Sari, Piran, and Narayan (1998) determined the synchrotron light-curves for these two regimes at high and low frequencies \( \nu > \nu_0 \) and \( \nu < \nu_0 \) respectively. If \( \varepsilon_\epsilon \ll 1 \), the evolution stays adiabatic throughout due to the little possibility of having radiative losses, \( \varepsilon_\epsilon \) being quite small. If during the fast cooling phase \( \varepsilon_\epsilon \) is somewhat less than unity, then only a fraction of the shock energy is lost to radiation. This situation is intermediate between the fully radiative and fully adiabatic cases discussed in Sari, Piran, and Narayan (1998).

Another point made by these authors is the possibility of defining the critical times \( t_c \) and \( t_m \) at which the break frequencies \( \nu_c \) and \( \nu_m \) cross the observed frequency \( \nu \). This is easily obtained when the time evolutions of \( \nu_c, \nu_m, F_{\nu,\text{max}} \) are known. These times \( t_c \) and \( t_m \) have been calculated in the two extreme regimes, adiabatic or radiative (see for instance Sari, Piran, & Narayan 1998, Piran, 2005). Figure 6.5 gives the synchrotron light-curves predicted for the two cases \( \nu > \nu_0 \) and \( \nu < \nu_0 \), ignoring self-absorption, the three critical times \( t_c, t_m, \) and \( t_0 \) (when \( \nu_c = \nu_m \)) are indicated.

More recently Wu et al. (2005b) have constructed afterglow light-curves that take into account the radiation losses of the blastwave. Its deceleration is considered in the two usual cases of a homogeneous interstellar medium and of a stellar wind environment. Their Figures 1 and 2 give the light-curves from high to low frequency. For instance they give the temporal decay index of the X-ray afterglow light-curve in the semi radiative phase \( \alpha = (3p - 2 - \varepsilon)/(4 - \varepsilon) \) in the ISM case and \( \alpha = (3p - 2 - (p - 2)\varepsilon)/2(2 - \varepsilon) \) in the stellar wind case, where \( \varepsilon \) is the radiation
Figure 6.5. Synchrotron light-curves ignoring self-absorption for two cases: high frequency ($\nu > \nu_0$) and low frequency ($\nu < \nu_0$). The flux variation with time is indicated for adiabatic evolution and the scaling in brackets is for radiative evolution which is restricted to $t < t_0$ when the evolution changes from radiative to adiabatic. $t_c$ and $t_m$ are the critical times when the break frequencies $\nu_c, \nu_m$ cross the observed frequency $\nu$. The spectral slopes during the time intervals B, C, D, F, G, H are given in Figure 5.4 (from Sari, Piran, & Narayan 1998).

efficiency. With $\varepsilon = 1/3$ and $p = 2.2–2.3$ the temporal index $\alpha_x$ of the X-ray afterglow light-curve seems consistent with the observed value $\sim 1.3$.

Finally it seemed interesting to introduce another way to study $F_\nu$. It was proposed by Sari, Piran, and Narayan (1998), to represent the observed emission
Table 6.1. Spectral ($\alpha$) and temporal ($\beta$) indices for fast cooling ($\nu_a < \nu_c < \nu_m$) into a constant density ISM (Piran 2005).

<table>
<thead>
<tr>
<th></th>
<th>$-\alpha$</th>
<th>$-\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu &lt; \nu_a$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$\nu_a &lt; \nu &lt; \nu_c$</td>
<td>$1/6$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$\nu_c &lt; \nu &lt; \nu_m$</td>
<td>$-1/4$</td>
<td>$-1/2$</td>
</tr>
<tr>
<td>$\nu_m &lt; \nu$</td>
<td>$-(3p-2)/4$</td>
<td>$-p/2 = (2\alpha - 1)/3$</td>
</tr>
</tbody>
</table>

Table 6.2. Spectral ($\alpha$) and temporal ($\beta$) indices for slow cooling ($\nu_a < \nu_m < \nu_c$) into a constant density ISM (Piran 2005).

<table>
<thead>
<tr>
<th></th>
<th>$-\alpha$</th>
<th>$-\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu &lt; \nu_a$</td>
<td>$1/2$</td>
<td>2</td>
</tr>
<tr>
<td>$\nu_a &lt; \nu &lt; \nu_m$</td>
<td>$1/2$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$\nu_m &lt; \nu &lt; \nu_c$</td>
<td>$-3(p-1)/4$</td>
<td>$-(p-1)/2 = 2\alpha/3$</td>
</tr>
<tr>
<td>$\nu_c &lt; \nu$</td>
<td>$-(3p-2)/4$</td>
<td>$-p/2 = (2\alpha - 1)/3$</td>
</tr>
</tbody>
</table>

by a series of power-law segments in time and frequency: $F_\nu \propto t^{-\alpha} \nu^{-\beta}$ separated by break frequencies at which the exponents of these power laws change. The identified breaks are $\nu_c$ the cooling frequency, $\nu_m$ the typical synchrotron frequency, and $\nu_{sa}$ the self-absorption frequency. The values of $\alpha$, the temporal decay index, and $\beta$, the spectral slope, are given in Tables 6.1 and 6.2 (from Piran (2005), his Tables I and II) for fast and slow-cooling and for $p > 2$. \(^1\)

Using $F_\nu \sim t^{-\alpha}$ in the adiabatic case (slow cooling), there are two possible slopes for the decaying part of the light-curve at a given time, hours to days after the GRB: $\alpha = \frac{3(p-2)}{4}$ and $\alpha = \frac{(3p-1)}{4}$. If the power-law index $p$ of the electron distribution is always the same and if the evolution is adiabatic, we expect to observe one of these two values of $\alpha$, differing by $1/4$. The X-ray afterglow appears in general to decay with $\alpha = 1.4$ after a few hours, but in some other cases (GRB 970228 and GRB 970508) the optical and X-ray afterglows follow $\alpha = 1.2$. The difference in $\alpha$ is consistent with $1/4$. These two possible values, which were observed with BeppoSAX hours and days after the GRB, were considered as a strong confirmation of the external shock model and of the adiabatic regime (Sari, Piran, & Narayan 1998).

\(^1\) The condition $p > 2$ is required so that the energy does not diverge at large $\gamma$; nevertheless some authors considered distributions with $2 > p > 1$ but with a maximal energy cut-off (e.g. Dai & Cheng 2001, Bhattacharya 2001).
We will see in Chapter 7 that Swift has revealed a much more complicated evolution of the X-ray afterglow, which required introducing additional complexity in the model.

Using now $F_\nu \sim \nu^{-\beta}$ in the adiabatic case (slow cooling), the two values of $\alpha$ correspond to $\beta = p/2$ and $\beta = (p - 1)/2$, respectively. Under these conditions for the adiabatic case $\alpha = (3\beta + 1)/2$ or $\alpha = 3\beta/2$. In a radiative evolution the relation would be $\alpha = (12\beta - 2)/7$.

These relations, called closure relations, have been investigated from a theoretical perspective by various authors (see Piran (2005) for a review, containing a total of six tables), for various conditions: adiabatic or radiative evolution, fast or slow cooling, ISM or wind environment. Tables 6.1 and 6.2 summarize the values of $\alpha$ and $\beta$ under various conditions. Looking at these different conditions and establishing such relations between $\alpha$ and $\beta$ for many GRBs and their afterglow appears a good way to approach a confirmation of the shock model, but also a way to rigorously specify the evolution of the afterglow. The measurement of $\alpha$ and $\beta$ should ideally be made at various wavelengths to measure the instantaneous position of the critical frequencies, but it is difficult to measure the evolution of the afterglow over a broad range of frequencies. Significant progress in this field has been accomplished with the fast alerts of Swift as discussed in Chapter 7.

### 6.5 THE STANDARD AFTERGLOW MODEL IN DETAIL

The standard afterglow model, which has been considered in the preceding parts, is based on the following approximations (Mészáros 2002).

(a) spherical outflow;
(b) line-of-sight scaling relations are assumed valid for the entire visible hemisphere;
(c) impulsive energy input $E_0$ and a single $\Gamma_0 = \eta = E_0/M_0c^2$;
(d) highly relativistic expansion in the adiabatic regime;
(e) homogeneous external medium: $n = n_0$;
(f) time-independent shock acceleration parameters $p, \varepsilon_e, \varepsilon_B$ (the energy index of the power-law electron spectrum, and the electron to proton and magnetic to proton energy ratios respectively);
(g) only the forward shock radiation is considered.

This model was quite successful in explaining many observations made by BeppoSAX in the first years after the famous GRB 970228. Wijers, Rees, and Mészáros (1997) gave a convincing comparison of the decline of the afterglow of GRB980228 at different wavelengths with the standard model (see Figure 3.4). Galama et al. (1998) explained the X-ray to radio spectrum of GRB 970508 measured 12.1 days after the event (Figure 3.7). Price et al. (2003) explained the X-ray to radio spectrum of GRB 030329 measured 0.5 days after the burst (Figure 4.4). These examples indicate that the approximations of the model are robust in a broad sense and over a range of timescales (Mészáros 2002).
Nevertheless the approximations listed above imply simplifications, which certainly do not exist in the real situation of most GRBs, so it is interesting to consider more complex conditions which make it possible to extend and complete the standard model. We will discuss below:

- The impact of the reverse shock crossing the ejecta.
- The beaming of the outflow within a jet of solid angle $\Omega_j < 4\pi$ which is an important ingredient to reduce the energetics of GRBs.
- The case of an external medium that is not homogeneous, with as an example the presence of a wind ejected by the progenitor prior to the burst.
- The departure from a simple impulsive injection (characterized by a delta or top-hat function with a single value for $E_0$ and $\Gamma_0$ in time) with the presence of refreshed shocks. A departure from the standard model can be for instance the injection of more energy with lower Lorentz factor at a later time, which still does not exceed the gamma-ray pulse duration. Such injections would drastically extend the relativistic regime by providing a late ‘energy refreshment’ to the blastwave on a timescale comparable to the afterglow timescale (Rees & Mészáros 1998).

Let us start the analysis of these situations with their consequences on the light-curves of the afterglow in various wavelength domains.

### 6.6 REVERSE SHOCK AND EARLY OPTICAL FLASHES

We have seen that afterglow observations are well described by the synchrotron emission from the ISM electrons accelerated by the forward shock. As can be seen in Figures 6.1 to 6.3, the interaction of an expanding relativistic shell with the ISM involves two shocks: the forward shock, which explains the afterglow, and the reverse shock (Mészáros & Rees 1994, 1997). The reverse shock is predicted to produce a strong optical flash and a radio flare (Mészáros & Rees 1997, Sari & Piran 1999b,c).

When it crosses the shell, the reverse shock heats up the shell’s matter and accelerates electrons, but it operates only once. Thus, unlike the forward shock emission that continues later at lower energy, the reverse shock emits a single burst. After the peak of the reverse shock no new electrons are injected and the shocked shell material cools adiabatically. This picture can be modified, allowing a long-lived reverse shock, if the central engine emits slowly moving material in which the reverse shock could propagate and survive for hours to days (Genet, Daigne, & Mochkovitch 2007, Úlmann & Beloborodov 2007; see Chapter 7).

Using standard assumptions, the forward-shocked ISM and the reverse-shocked shell carry comparable amounts of energy (Sari & Piran 1999b). However, the typical temperature of the shocked ejecta is much lower than that of the shocked ISM since the mass density of the ejecta is higher and the shock weaker. Consequently, the typical frequency of the synchrotron emission from the reverse shock is much lower than that from the shocked interstellar medium and the reverse shock is expected to emit significant radiation at lower frequencies, particularly in the optical band. The
forward and reverse shocks define four distinct regions: the unshocked ejecta, the shocked ejecta, the shocked wind or ISM, and the unshocked wind or ISM (Figure 6.3). The circum-burst medium can be the interstellar medium (ISM) or a wind. Kobayashi et al. have considered these two cases for both thin and thick shells (Kobayashi 2000, Kobayashi & Sari 2000, Kobayashi & Zhang 2003a). Chevalier & Li (2000) have considered the thick shell case for the wind model.

Using the jump conditions for the shock and the equality of pressure and velocity along the contact discontinuity (Figure 6.3), the Lorentz factor $\Gamma$, the pressure $p$, and the number density $n$ in the shocked regions can be estimated as functions of $n_1$, $n_4$, and $\Gamma$ the Lorentz factor of the relativistic shell (Blandford & McKee 1976). There are two limiting cases:

- If the shell density is high ($n_4 \gg \Gamma^2 n_1$), the reverse shock is Newtonian, and $\Gamma_3$, the Lorentz factor of the shocked shell material, is almost unity in the frame of unshocked shell material. The reverse shock is too weak to slow down the shell effectively ($\Gamma_3 \sim \Gamma$). The Lorentz factor of the shocked shell material is almost constant while the shock propagates through the shell.

- If the density is low ($n_4 \ll \Gamma^2 n_1$) the reverse shock is relativistic and it considerably decelerates the shell material. This situation can be explained as follows: when a shell is ejected from a source, it has a high density so the reverse shock is initially Newtonian. As the shell expands, its density decreases and the reverse shock may become relativistic while it crosses the shell.

Kobayashi (2000) considered the two cases, determining $\Gamma_3$, $n_3$, $p_3$, and $N_e$ ($N_e$ being the number of shocked electrons) and their dependence versus the observer time $t$ for the reverse shock. Considering the synchrotron emission of the electrons, light-curves are obtained for the thick and thin shell cases which correspond to long and short bursts respectively. Kobayashi (2000) applied his model to the prompt optical emission from GRB 990123 (Akerlof et al. 1999; see Section 3.9 and Figure 3.15) and he concluded that the optical flash is well described by the reverse shock emission in the case of a homogeneous circum-burst medium, while the wind model appeared inconsistent with the observations (Kobayashi & Zhang 2003a). This conclusion is also shared by Sari and Piran (1999c) and Mészáros and Rees (1999). The lack of correlation between the prompt optical and gamma-ray emissions for this burst suggests that different processes produced the GRB and the optical flash. Other analyses of the optical flash of GRB 990123 have also invoked the reverse shock emission as the source of this flash (Wang et al. 2000, Panaitescu & Kumar 2001b, Fan et al. 2002, Soderberg & Ramirez-Ruiz 2002, Zhang, Kobayashi, & Mészáros 2003, Nakar & Piran 2005).

The explanation of optical flashes by the reverse shock has led to some estimates of the initial value of the Lorentz factor $\Gamma_0$. For GRB 990123 the possible values for $\Gamma_0$ are scattered between 200–300 and 900–1200 (Sari & Piran 1999c, Zhang, Kobayashi, & Mészáros 2003, Soderberg & Ramirez-Ruiz 2002, Wang, Dai, & Lu 2000a). For GRB 990123, assuming the optical flash is due to the reverse shock, Panaitescu & Kumar (2001b) gave a rough estimate of $\Gamma_0 = 1400 \pm 700$. These values
between a few hundred and a thousand are not precise enough to constrain the baryonic loading of the fireball, nor to improve our understanding of the nature of GRBs, specifically the mechanism and location of the GRB prompt emission. In addition to the optical flash, GRB 990123 was followed by a radio flare (Kulkarni et al. 1999b). While radio afterglows usually rise on timescales of weeks or more, in the case of GRB 990123 the radio flare peaked at 1 day with a fast rise and decay. Sari and Piran (1999c) have interpreted this flare as the emission from particles in the ejecta previously shocked by the reverse shock. The shocked ejecta initially radiates in the optical band. As the ejecta expends the temperature of the ejecta is lower and the electrons cool due to this adiabatic expansion. The emission frequency and the flux drop quickly. Eventually the emission comes to the radio band and even lower. Therefore this radio flare appears as the late-time equivalent of the early optical flash (Sari & Piran 1999c, Kobayashi & Sari 2000). The reverse shock has also been invoked to explain the early radio afterglow of the unusual GRB 991216 (Frail et al. 2000), whose X-ray and optical afterglows were fairly typical (Halpern et al. 2000).

6.6.1 Some examples

After the first observation of an optical flash associated with GRB 990123 a few additional examples of such simultaneous or almost simultaneous optical emission have been reported thanks to the prompt alerts of the HETE-2 satellite (the observations of Swift GRBs are discussed in Chapter 7). Let us consider some examples.

**GRB 021211**

The wide-field 1.2-m Palomar Oschin telescope equipped with the Near Earth Asteroid Tracking camera (Pravdo et al. 1999) observed this burst 20 min after the GRB trigger time (Fox & Price 2002). The Katzman Automatic Imaging Telescope (KAIT) at Lick Observatory started a pre-programmed sequence of images at 171 s (Li et al. 2002). Raptor obtained an image at \( t = 65 \) s (Wozniak et al. 2002), and Superlotis observations began at \( t = 143 \) s (Park, Williams, & Barthelmy 2002). Fox et al. (2003a) reported the early-time optical, near-infrared, and radio follow-up observations of the afterglow. The early afterglow was roughly three magnitudes fainter than the afterglow of GRB 990123 at similar epochs and fainter than almost all known afterglows at 1 day after the GRB (see Figure 6.6, which also shows a two-component flare + afterglow fit for the early optical emission). Infrared observations indicated that this faintness could not be attributed to extinction. The early behavior of this burst was quite similar to the early light-curve of GRB 990123 (Chornock et al. 2002; see also Figure 3 of Li et al. 2002). Its steep declining flash followed: \( f \propto t^{-\alpha} \) with \( \alpha \approx 1.6 \). As in the case of GRB 990123 the observed optical flash was supposed to result from a reverse shock propagating into the GRB ejecta (Fox et al. 2003a). Kumar and Panaitescu (2003) noted that the synchrotron radiation from the forward shock could account for the gamma-ray emission of this burst (one single peak about 2 s long) without the need to invoke internal shocks. They also considered that the
Figure 6.6. The optical light-curve of GRB 021211 starting ~90 seconds after the burst. The thick solid line shows a two component fit (flare + afterglow) of the early optical emission (from Fox et al. 2003a). The fit to the data is done for $t < 0.1$ days, the dotted line at $t > 0.1$ days is only illustrative. Data for other afterglows are also reported for comparison.

Early optical emission of this GRB was presumably coming from the reverse shock heating of the ejecta. They suggested that the magnetic field in the reverse shock and in the early forward shock is a frozen-in field originating in the explosion, and that most of the energy in the explosion was initially stored in the magnetic field. Therefore, as noted also by Zhang, Kobayashi, & Mészáros (2003), there was evidence that the central engine was strongly magnetized as for GRB 990123. They finally suggested that the role of the magnetic field could be important to trigger the prompt emission and to collimate the GRB jets. However, as there was a reverse shock, the fireball could not be completely Poynting flux–dominated.
Other early optical emissions have been reported, for instance:

- **GRB 021004.** This GRB was detected by HETE-2 and promptly localized, allowing an early follow-up of the afterglow. Fox et al. (2003b) detected an optical transient 193 seconds after the trigger, and Torii, Kato, and Yamaoka, (2002) observed the error box $\sim 210$ s after the trigger. The peculiarity of the early afterglow light-curve is the presence of a major rebrightening around 0.1 days and of short-term variability around $\sim 1$ day. The slope of the afterglow is small ($\sim t^{-0.7}$) for the earliest observations (9–17 min) and it decays with a more conventional slope ($t^{-1.05}$) after 0.1 day. Kobayashi and Zhang (2003b) have fit the optical light-curve with the contributions of the reverse and forward shock emissions (Figure 6.7). They propose that the optical emission is dominated by the reverse shock at early times, explaining the peculiar temporal evolution of the

![Figure 6.7. Optical light-curve of GRB 021004. The references for the different measurements are given in Kobayashi & Zhang (2003b). The rising part of the afterglow was obtained thanks to the rapid alert of HETE-2 and early optical follow-up observations. The optical light-curve model presented by these authors includes the forward shock emission (dash-dotted line, with an initial rise followed by a decay when the typical synchrotron frequency, $\nu_\text{m}$, crosses the observation band), the reverse shock emission, which can be identified in the very early afterglow (dashed line) and the total flux (thick solid line). This model is computed for $p = 2.4$, $\varepsilon_B = 3.0 \times 10^{-3}$, $\varepsilon_e = 0.28$, $n = 0.45$ proton cm$^{-3}$, $\eta = 120$, and $E = 5.6 \times 10^{52}$ erg. Another model (thin solid line) is given for slightly different parameters but the same total energy and circum-burst density (from Kobayashi & Zhang 2003b).](image)
early afterglow. Other explanations have been proposed: Lazzati et al. (2002), for instance, showed that the bumps in the afterglow light-curve could be due to the interaction of the fireball with clumps of matter embedded in a uniform environment (see also Section 4.2.1). The observed temporal features could also be modeled with refreshed shocks (Björnsson, Gudmundsson, & Jóhannesson 2004).

- **GRB 030418.** An early optical observation of this burst was made with the ROTSE-IIIa telescope in Australia 211 s after the start of the burst and only 76 s after its end (Rykoff et al. 2004). The light-curve of the afterglow shows a rising part during 600 s followed by a slowly varying plateau for 1400 s and then a classical power-law decay ($f_\nu \propto t^{-\alpha}$) with slope $\alpha \sim 1.36$.

- The early afterglow of GRB 030723 imaged with the ROTSE-IIIb telescope (in Texas, USA) showed a similar behavior but two magnitudes dimmer. These two afterglows are different from those of GRB 990123 and GRB 021211 as they do not show a bright early optical flash, which could be due to the reverse shock. Rykoff et al. (2004) suggested that the slowly rising part of the afterglow could be attributed to a strong extinction in the local circum-burst medium. The extinction slowly decreases as the afterglow light is emitted farther and farther from the GRB.

With Swift the number of GRBs with very early optical afterglow increased significantly, as discussed in Chapter 7.

### 6.6.2 Alternative explanations for early optical flashes

Other possibilities have been proposed for the origin of early optical flashes. Mészáros and Rees (1997) considered for instance that such optical flashes might result from internal shocks even if optical flashes from internal shocks are nearly two orders of magnitude weaker than those due to the reverse shock (for the same total energy). They calculated that, with beaming factors of $\sim 10^{-2}$, it is possible to have flashes as bright as 9th magnitude for $z \sim 1$ (see also Mészáros & Rees 1999, Kumar & Piran 2000b).

With respect to this possibility, the case of the early optical emission of GRB 041219A is interesting. The optical detection started very early—8 s after the receipt of the INTEGRAL alert—with the Raptor telescopes (Vestrand et al. 2002). In this burst the prompt optical emission appears variable and correlated with the prompt gamma-rays. Figure 6.8 (Vestrand et al. 2005) shows the correlation of the optical flash with the GRB. What is also interesting in this figure is the remarkable similarity between the gamma-ray light-curves of GRB 041219A and GRB 990123, which is not reflected in their optical emission, for instance the optical emission was low during the first gamma-ray pulse of GRB 990123, in contrast to GRB 041219A. In this last burst, the fast rise of the optical emission, simultaneous with the dominant gamma-ray pulse, and its overall correlation with the prompt gamma-ray emission suggest that the emissions in the two energy bands were generated by a common mechanism. Figure 3 of Vestrand et al. (2005) shows that the prompt optical flux is reasonably well explained by assuming a constant optical to $\gamma$-ray flux ratio. The authors
Figure 6.8. Comparison of prompt gamma-ray and prompt optical-light curves of GRB 041219A and the famous GRB 990123. In 2004 these were the only two cases where prompt optical emission had been detected. The gamma-ray light-curves are similar whereas the two optical curves are different: the optical maximum occurs early for GRB 041219A and after the main gamma-ray emission for GRB 990123. Moreover, it is interesting to note a correlation between the gamma ray and the optical emissions of GRB 041219A, while this correlation is completely absent for GRB 990123 (from Vestrand et al. 2005).

considered the possibility that the prompt optical emission is the low-energy tail of the synchrotron emission generated by internal shocks in the GRB outflow. In any case, in this burst the reverse shock emission must be suppressed and/or delayed (Vestrand et al. 2005).

A third mechanism has been proposed: the early optical emission could be produced by the forward shock at its early stage when it propagates into the pre-accelerated and pair-loaded environment (Beloborodov 2002; see also Mészáros, Ramirez-Ruiz, & Rees 2001).
6.6.3 Suppressing the reverse shock optical emission?

After the detection of GRB 990123 and its bright optical emission, Kobayashi (2000) studied GRB 981121 and GRB 981223, which were observed by ROTSE (Akerlof et al. 2000) during the prompt GRB phase, but not detected in the optical domain. For these two GRBs, if they assumed a ratio of the optical to gamma-ray fluence similar to GRB 990123, the optical emissions should be more than two magnitudes over the ROTSE detection thresholds. This argument was not very strong, however, because these two GRBs were not localized precisely, preventing the detection of their late optical afterglows, and consequently they could just have been heavily absorbed dark GRBs. Even if these two GRBs were not dark, there is no reason to have a constant optical to gamma-ray fluence ratio if the optical flash is due to the reverse shock since the GRB is due to internal shocks. Other suggestions might explain the lack of bright optical emission from the reverse shock: Sari and Piran (1999b) noted that the calculations for the reverse-shock emission assumed that the shell was made out of baryonic material. If instead it is magnetically dominated, the energy in the rest mass of baryons will be negligible and a considerably lower emission is expected from the reverse shock. Moreover the preceding predictions are based on the fact that the reverse and forward shocks carry nearly the same amount of thermal energy. If the shell is initially very thin and does not spread so that its thickness is kept significantly below $R/\Gamma^2$ (the forward shock compresses the fluid ahead of it by a factor $\Gamma^2$) the reverse shock will be Newtonian, it will contain only a small fraction of the energy (Sari & Piran 1995) and the emission will be reduced accordingly. This is, however, a very special case since when the shells are accelerated hydrodynamically the spreading of the shell is natural. Finally, the reverse-shock energy can be radiated at non-optical frequencies.

Over time, it appeared that GRB 990123 was really unique with respect to its optical brightness. Despite a growing number of prompt observations, no GRB was found which was so bright in the optical (until GRB 080319B). After 2004 new analyses of the reverse-shock emission were proposed with complementary conclusions. They contributed to provide answers to the following question: why are there so few bright early optical flashes?

(a) Nakar and Piran (2004), for instance, presented a new fingerprint of the reverse-shock emission in addition to the well-known $t^{-2}$ optical decay. The observation of this signature, the consistency between peak time and flux at optical and radio wavelengths would indicate that the reverse shock dominates the emission during the early afterglow and that the circum-burst medium is ISM-like. These new calculations allowed Nakar and Piran (2004) to conclude that previous calculations overestimated the intensity of the optical flash by order(s) of magnitude. In fact an optical flash brighter than 15 mag is expected in some GRBs, but not all. A typical value $R \sim 17–19$ is expected when the reverse shock is mildly relativistic ($\xi \sim 1$, where $\xi$ is a parameter describing the nature of the shock—relativistic vs Newtonian—introduced in Section 6.2). When the reverse shock is relativistic ($\xi \ll 1$) most of the emission is released in the optical domain. When $\xi$ increases,
the emission is shifted toward lower energies but it does not reach the radio. So, only a fraction of the flashes are expected to be bright enough for detection at optical wavelengths.

(b) Another interesting point concerns the role of the pairs in the ejecta. This question was considered by McMahon, Kumar, and Piran in 2006, with the idea of seeing if early afterglow observations could be used to determine the pair content of GRBs. They showed that the ratio of optical to IR flux depends very weakly on the number of pairs in the ejecta (see also Li et al. 2003), while the flux ratio of the optical to radio bands can be used because it depends on the number of pairs per proton. In fact pair-enrichment of the ejecta causes the reverse-shock emission to be fainter than that expected from completely baryonic ejecta. Strong magnetic fields in the ejecta have the same effect: they weaken the reverse shock (see also Zhang & Kobayashi 2005). McMahon, Kumar, and Piran (2006) found that reverse-shock emission is very weak over a large parameter space. Moreover, the optical flux from the forward shock is often brighter than the reverse shock at deceleration. The authors concluded that these facts could explain the paucity of prompt optical flashes with a rapidly declining light-curve at early time (McMahon, Kumar, & Piran 2006). McMahon, Kumar, and Panaitescu (2004) have also investigated the ways to have a weak reverse-shock emission. If the cooling frequency in the reverse shock is below the optical band, the observed flux decays rapidly ($\sim t^{-3}$) and it can be unobservable after few deceleration times. It is also possible that the deceleration time is much longer than the duration of the burst in which case a weak reverse shock is expected (McMahon, Kumar, & Panaitescu 2004).

(c) The degree of magnetization of the ejecta is another important issue because, as we have seen, evidence exists that some GRB ejecta are magnetized. For instance, Zhang, Kobayashi, and Mészáros (2003) showed that the interpretation of the early afterglow of GRB 990123 and GRB 021211 requires a stronger magnetic field in the reverse-shock region than in the forward shock region. In fact the characteristics of the reverse shock are strongly influenced by the degree of magnetization of the ejecta. Zhang and Kobayashi (2005) considered this problem and studied the reverse-shock emission properties of arbitrarily magnetized ejecta assuming a constant density of the circum-burst medium. They characterized the degree of magnetization using the parameter $\sigma$, which is the ratio of the electromagnetic energy to the kinetic energy flux (see their relation 7). When magnetic fields are introduced through the equipartition parameter with values of $\varepsilon_B$ ranging from 0.001 to 0.1, the ejecta are in the $\sigma \rightarrow 0$ regime. When $\sigma \rightarrow \infty$ the Poynting flux dominates the flow and the GRB emission is dominated by magnetic dissipation processes. Zhang and Kobayashi (2005) computed the early optical light-curves expected for a wide range of $\sigma$ values: $\sigma = 0, 0.001, 0.01, 0.1, 1, 10, \text{and} 100$. Their Figure 7 gives the early afterglow for all these values of $\sigma$, showing that early afterglow data might be used to diagnose the magnetic content of GRB ejecta. Among the various conclusions of their work, novel features of the early light-curves are pointed out. As $\sigma$ increases, the peak flux of the reverse shock starts to increase rapidly,
reaching a maximum around $\sigma \sim 0.1$–1, and then decreases when $\sigma > 1$. When $\sigma$ becomes larger than several tens to several hundreds the conditions to have a reverse shock are no longer met and no reverse-shock component can be expected in the early afterglow light-curves. The Swift dark burst GRB 050319A (Roming et al. 2006) could be in this category. Zhang and Kobayashi suggested that the observed diversity of early afterglows can be explained if $\sigma$ is allowed to vary over a wide range, from 0.01 to 100.

(d) Another parameter impacting the properties of the reverse shock is the density of environment. Kobayashi, Mészáros, and Zhang (2004) reported that a more complex optical/IR light-curve is expected in a dense environment. The reverse-shock spectrum presents a prominent bump, especially if the burst ejecta are highly magnetized (their Figure 1). The schematic optical light-curve for a synchrotron-dominated fireball in a dense environment or wind is given in their Figure 2 (Figure 6.9). The observation of early optical/IR and/or X-ray emission attributable to reverse-shock emission can thus be used also to constrain the GRB environment. In fact prompt optical flashes from the reverse shock in a dense environment contain both spectral and temporal characteristic signatures. Moreover, in such winds or a dense ISM the spectra and light-curves can also give constraints on the strength of the magnetic field in the ejecta.

In conclusion there are many reasons to have a weak reverse shock and a lack of bright prompt optical flashes. This is in agreement with the limited number of observed optical flashes in spite of the very rapid alerts after the GRB detection and of the quality and the number of robotic telescopes around the world. Even though the number of events with early optical emission attributable to the reverse

![Figure 6.9. Schematic optical light-curve for a synchrotron dominated fireball in a dense environment, such as a wind or a dense interstellar medium. The reverse shock emission (solid line) is characterized by a bump, which is more prominent if the burst ejecta are magnetized. The dashed line shows the forward-shock emission. Timescales are indicative (from Kobayashi, Mészáros & Zhang 2004).](image-url)
shock is today limited, the scientific potential of this emission (light-curve and spectra) is quite important, especially to understand the degree of magnetization of the ejecta and the properties of the circum-burst environment. Significant progress in this field has been made with Swift bursts, as we will see in Chapter 7.

6.7 NON-SPHERICAL RELATIVISTIC EJECTA: JETS

Some pieces of evidence suggest that GRB outflows are collimated into narrow jets. An indirect but quite strong argument comes from the very high values of the energy output in gamma-rays if isotropic emission is assumed. The isotropic equivalent energy, the energy release inferred when assuming isotropic emission, approaches and sometimes exceeds $M_{\odot}c^2$, the binding energy of the Sun. It is therefore difficult to produce such events in models involving stellar mass progenitors. On the other hand, if GRBs radiate into a narrow jet, the true energy output in gamma-rays can be greatly reduced, by one to three orders of magnitude.

A more direct line of evidence in favor of collimated outflows comes from long-term radio observations: at late times (several months or years) when the outflow becomes sub-relativistic, it radiates isotropically at radio wavelengths. Consequently, late radio observations permit measurement of the total kinetic energy of the ejecta (e.g. Waxman, Kulkarni, & Frail 1998, Frail, Waxman, & Kulkarni 2000). The few measures performed to date have led to energies clustering around a few $\times 10^{51}$ erg, well below the numbers found assuming isotropic emission.

6.7.1 Jet breaks

Additional indirect evidence is provided by the achromatic breaks seen in the afterglow light-curves of several GRBs (Rhoads 1997, Sari, Piran, & Halpern 1999, and Section 3.4). The uniform (or top hat) jet model is the first type of jet that was introduced (Rhoads 1997, 1999, Panaitescu & Mészáros 1999, Sari, Piran, & Halpern 1999, Kumar & Panaitescu 2000b, Moderski, Sikora, & Bulik 2000, Granot et al. 2002). In this “uniform” jet the initial energy per solid angle $\epsilon$ and the Lorentz factor $\Gamma$ are uniform within some finite half opening angle $\theta_j$ and sharply drop outside $\theta_j$; such a situation corresponds to non-spherical relativistic ejecta.

Due to special relativistic effects, the radiation from a blob of matter with a bulk Lorentz factor $\Gamma$ is beamed within a beaming angle $\Gamma^{-1}$. As matter is initially ultra-relativistic, $\Gamma^{-1}$ is quite small and observers only see a tiny fraction of the jet and they cannot determine if the ejecta are beamed or isotropic. During the deceleration phase $\Gamma$ decreases until $\Gamma^{-1}$ reaches $\theta_j$ at a distance $R_\theta$. This happens much later than the GRB phase, in the afterglow phase. At about the same time the jet starts to expand laterally. At that time ($\Gamma \sim \theta_j^{-1}$), the hydrodynamic evolution of the source changes. Initially, when $\Gamma \gg \theta_j^{-1}$, the motion is almost conical. There is not enough time in the blastwave rest-frame for the matter to be affected by the non-spherical geometry and the blastwave behaves as if it were a part of a sphere (Piran 2005). When $\Gamma$ falls below $\theta_j^{-1}$ the jet spreads in the lateral direction. Rhoads (1997, 1999) found that the jet will
expand sideways at the sound speed of shocked relativistic material \( c_s = c/\sqrt{3} \). In this regime \( \Gamma \) decreases exponentially with radius \( r \), and not as a power law. The flux at the fixed frequency above \( \nu_m \) drops as \( F_\nu = F_{\nu m}(\nu/\nu_m)^{-(p-1)/2} \) or \( F_\nu \propto t^{-p} \). For \( p \sim 2.5 \) the decay index is much larger than that for a spherical afterglow \( t^{-(p-1)/4} \sim t^{-1.5} \) (Rhoads 1997, 1999, Sari, Piran, & Narayan 1998).

The beaming of the ejecta results in the existence of two breaks. The first is geometric due to the lack of emission from outside the jet, which becomes detectable when \( \Gamma \sim \theta_j^{-1} \). The second is hydrodynamic due to the sideways expansion of the jet. Panaitescu & Mészáros (1999) found that the ratio of the times of these two breaks is analytically overestimated at ~50 while numerical simulations lead to a substantially smaller value, <10. This ratio must still be reduced due to the curvature and thickness of the shell. Therefore, the authors concluded that the two breaks overlap. Like Rhoads (1999), Panaitescu and Mészáros (1999) assumed that the speed of the lateral expansion is the sound speed, \( c_s = c/\sqrt{3} \) or smaller. Sari, Piran, and Halpern (1999) proposed instead that the speed of the lateral expansion is the speed of light. They also find that the two breaks are merged, but the time of these breaks is very different from those given by Rhoads (1999). For instance, it differs by a factor of 20 in time from the expression given by Rhoads (1999). This has important consequences as it may result in a factor of 3 in the estimation of the opening angle of the jet (see below).

Another calculation of afterglow light-curves was made by Wei and Jin (2003) considering the possible large range of viewing angles \( (\theta_v) \) from the jet axis. For uniform jets the afterglow light-curves present two breaks: the first one when \( \gamma \sim (\theta_j - \theta_v)^{-1} \), where \( \theta_j \) is the jet opening angle, and the second one when \( \gamma \sim (\theta_j + \theta_v)^{-1} \). The calculations for the case of non-uniform jet are more complicated, and they show a clear flattening of the light-curve followed by a more prominent break. The authors suggested that the light-curves could be used to provide insight into the geometry of the jet.

Finally some authors have suggested that jet breaks could be essentially absent or very weak. Based on numerical studies, Moderski, Sikora, and Bulik (2000) concluded that the light-curve break is weaker and much smoother than predicted analytically. A steepening of the light-curve occurs but the slope change is small (the slope after the break is about 2.0 instead of \( \alpha = 2.4 \)), and it is extended over more than two decades of observer time (Moderski, Sikora, & Bulik 2000). This is also the conclusion of Wei and Lu (2000) who find that a significant break is only present when the ejecta have a very small opening angle, \( \theta_j < 0.1 \). Huang et al. (2000) reported that no obvious break is expected during the relativistic phase of the ejecta. However, an obvious break is expected at the transition from the relativistic phase to the Newtonian phase (a sharp decay in the optical afterglow), after 1 day or more. Curiously, Panaitescu and Mészáros (1999) noted that sideways expansion might slow down the flux decay rather than producing a faster one: even if the flux from the head-on part of the remnant decreases faster in this regime, this is compensated for by the increased emission measured from sweeping up external matter over a larger angle, and by the fact that this extra radiation which is emitted at larger angles arrives later (Mészáros, Rees, & Wijers 1999). According to the simulations, this is, however, a marginal situation.
2D hydrodynamic simulations (Granot et al. 2001) show that the jet does not spread sideways as much as predicted by simple analytical models. The shock front is egg-shaped and far from being spherical, and most of the radiation originates at the front of the jet (Figure 6.10). The determination of the surface of equal arrival times at the observer from a thin relativistic shell can be found in Rees (1966), Waxman (1997a), Sari (1998), Panaitescu and Mészáros (1998b), and Granot, Piran, and Sari (1999a,b). More recently 3D relativistic hydrodynamics GRB simulations (Cannizzo, Gehrels, & Vishniac 2004) confirmed the steepening of the light-curve roughly when the edge of the jet starts to be seen ($\Gamma^{-1} \sim \theta_j$). The significant differences between analytical models and 2D or 3D simulations indicate that great care has to be taken in the choice of the hydrodynamical model because the calculated light-curves depend strongly on the hydrodynamics of the jet.

The question of the lateral spreading of the jet is very important. Granot (2007) emphasized the fact that lateral expansion of the jet is very modest as long as the head of the jet is relativistic, in contrast to the very rapid lateral expansion predicted by semi-analytic models. This implies that the lateral spreading cannot play an important role in the jet break and the predominant cause of the break is the missing contribution from outside of the jet once its edge becomes visible (Granot 2007). Under these conditions the jet half opening angle remains roughly constant and the
asymptotic change in the temporal decay index \( \alpha (F_v \sim t^{-\alpha}) \) is only \( \Delta \alpha \sim 3/4 \) for a uniform external medium and \( \Delta \alpha = 1/2 \) for a wind (Figure 6.11). These predictions can be compared to the observed values reported by Zeh, Klose, and Kann (2006) for 16 afterglows observed in the pre-Swift era. They found that \( \Delta \alpha \), the difference in the decay slopes before and after the break, shows a maximum around 0.8 and a long tail toward higher values (see their Figure 3). In addition, they do not confirm the bimodality of the distribution of \( \Delta \alpha \) found by Panaitescu (2005b). These values of \( \Delta \alpha \) can be explained by the overshoot in the value of \( \alpha \) just after the time of the jet break and by the fact that in general the measurement of the post jet break temporal decay index \( \alpha_2 \) is made a short time after the break (Figure 6.11).

The properties of the afterglow depend on the medium surrounding the source. In a wind-type medium, the afterglow is expected to decrease more rapidly (Chevalier 
& Li 1999). For beamed GRBs, the magnitude and time of the breaks depend critically on the surrounding medium (wind vs. ISM). Various authors have studied the properties of the jet breaks when the jet propagates into a wind rather than into a homogeneous interstellar stellar medium. Kumar and Panaitescu (2000b) showed that afterglows from GRB remnants going off in a wind-type medium, with density decreasing as \( r^{-2} \) show little evidence for light-curve steepening when \( \Gamma \theta_j \sim 1 \), even when the jet is highly collimated. Panaitescu and Kumar (2000b) looked at the jet break transition in a wind profile and found that it appears after a much longer time. Therefore the observation of the break is much more problematic in this case since the flux of the afterglow is already very faint when the break occurs. This might explain the lack of breaks in the afterglow of several GRBs. Among them GRB 980326 and GRB980519 were found by Chevalier and Li (1999) to be good candidates for a wind model. This analysis seems to prove that, for different reasons, the jet may be not observed even if the ejecta are collimated. This point will be discussed in more detail in Chapter 7, in light of the Swift observations. Frail et al. (2000b) showed that radio observations can be decisive to interpret the fast decline of afterglows as due to a jet or to a blast wave propagating into a medium whose density is shaped by a wind. Their Figure 2 clearly shows the very different radio flux expected 1 month after the GRB. For instance, the radio observations of GRB 980519 at 8.46 GHz are clearly in favor of a wind model rather than a homogeneous model; nevertheless too few radio observations are available to really choose the best model for many GRBs.

After the jet break the temporal decay index is predicted to be \( \alpha = p \). This result assumes significant sideways expansion and it does not depend on whether the observed X-ray band is below or above \( \nu_c \) and on whether the medium is an ISM or a stellar wind. It also assumes that \( p \), the electron spectral index, is >2. A complete census of the values of \( \alpha \) for various afterglow conditions and frequency domains (with respect to \( \nu_c \) and \( \nu_m \)) is given in Table 6.3 (reproducing Table 2 of Zhang et al. 2006).

Even though we have emphasized the predictions and phenomenology of jet breaks in the light-curves of GRB afterglows, it is worth mentioning that (i) temporal breaks can have other origins (e.g. cooling breaks, see Section 6.4), and (ii) jetted GRBs have another important consequence which is the existence of ‘orphan after-
Figure 6.11. Shape of the jet break characterized by the evolution of the temporal decay index $\alpha$ as a function of the observed time (in days). In the upper panel the evolution of $\alpha$ across the jet break is given for two different environments, and for two different models in the lower panel (from Granot 2007). The upper panel (a) gives the evolution of $\alpha$ using a semi-analytic model with no lateral spreading (Granot 2005) for a uniform external density profile ($n(R) = \text{cst}$, $n_{\text{ext}} = 1 \text{ cm}^{-3}$) and for a wind density profile ($n(R) \propto R^{-3}$, $A_* = 1$). The lower panel (b) gives the evolution of $\alpha$ for the same uniform density profile calculated with a hydrodynamic simulation and with the aforementioned semi-analytic model (Granot et al. 2001, Granot 2005). The overshoot is clearly seen when hydrodynamic simulations are used. In both panels (a & b) the dashed lines show the asymptotic values of $\alpha$ before and after the jet break. Just after the break the values of $\alpha$ are far from asymptotic and $\Delta \alpha$ inferred from early observations after the break can be significantly larger than its asymptotic value.
Table 6.3. Temporal index \( \alpha \) and spectral index \( \beta \) in various afterglow models. The values in the table follow the convention \( F_\nu \propto t^{-\alpha} \nu^{-\beta} \). GRB models are characterized by the position of the observing frequency, \( \nu \), by the type of surrounding medium (ISM or wind), and by the cooling regime (fast cooling—FC, or slow cooling—SC). Models with energy injection are characterized by the parameter \( q \), which specifies the energy injection rate by the central engine (\( L(t) = L_0(t/t_\text{b})^{-q} \)). The temporal indices with energy injection are valid only for \( q < 1 \), and they reduce to the standard case (without energy injection) when \( q = 1 \). For \( q > 1 \) the expressions are no longer valid, and the standard model applies. An injection case due to pulsar spindown corresponds to \( q = 0 \) (Dai & Lu 1998a, Zhang & Mészáros 2001). Recent Swift XRT data are generally consistent with \( q \sim 0.5 \). The numerical values quoted in parentheses are for \( p = 2.3 \) and \( q = 0.5 \) (from Zhang et al. 2006).

<table>
<thead>
<tr>
<th>GRB models</th>
<th>( \beta )</th>
<th>( \alpha )</th>
<th>( \alpha(\beta) )</th>
<th>( \alpha )</th>
<th>( \alpha(\beta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISM, SC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu &lt; \nu_m )</td>
<td>(-\frac{1}{3})</td>
<td>(-\frac{1}{3})</td>
<td>(-\frac{1}{3})</td>
<td>(\frac{5q - 8}{6}(-0.9))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu_m &lt; \nu &lt; \nu_c )</td>
<td>(p \frac{1 - p}{2} (0.65))</td>
<td>(3(p - 1))</td>
<td>(\frac{1}{2})</td>
<td>(\frac{2}{4}(-1.0))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu &gt; \nu_c )</td>
<td>(p \frac{1}{2} (1.15))</td>
<td>(3p - 2)</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>ISM, FC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu &lt; \nu_c )</td>
<td>(-\frac{1}{3})</td>
<td>(-\frac{1}{6})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{7q - 8}{6}(-0.8))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu_c &lt; \nu &lt; \nu_m )</td>
<td>(\frac{1}{2})</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{3q - 2}{4}(-0.1))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu &gt; \nu_m )</td>
<td>(p \frac{1 - p}{2} (1.15))</td>
<td>(3p - 2)</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>Wind, SC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu &lt; \nu_m )</td>
<td>(-\frac{1}{3})</td>
<td>(0)</td>
<td>(\frac{3\beta + 1}{2})</td>
<td>(\frac{q - 1}{3}(-0.2))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu_m &lt; \nu &lt; \nu_c )</td>
<td>(p \frac{1 - p}{2} (0.65))</td>
<td>(3(p - 1))</td>
<td>(\frac{1}{4})</td>
<td>(\frac{2p - 2 + (p + 1)q}{4}(1.1))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu &gt; \nu_c )</td>
<td>(p \frac{1}{2} (1.15))</td>
<td>(3p - 2)</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>Wind, FC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu &lt; \nu_c )</td>
<td>(-\frac{1}{3})</td>
<td>(\frac{2}{3})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{1 + q}{3}(0.5))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu_c &lt; \nu &lt; \nu_m )</td>
<td>(\frac{1}{2})</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{3q - 2}{4}(-0.1))</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
<tr>
<td>( \nu &gt; \nu_m )</td>
<td>(p \frac{1}{2} (1.15))</td>
<td>(3p - 2)</td>
<td>(\frac{1}{4})</td>
<td>(\frac{1}{2})</td>
<td>(\frac{q - 2}{2} + \frac{(2 + q)\beta}{2})</td>
</tr>
</tbody>
</table>

glows’ discussed in Section 6.7.3. An ‘off-axis’ observer who does not see the original jet will be able to see an ‘orphan afterglow’, that is, an afterglow without a GRB. Being outside the opening angle \( \theta_j \) of the jet, the observer cannot see the GRB but he will be illuminated by the afterglow when \( \Gamma^{-1} \) becomes larger than the angle of the jet
axis with his line of sight. This is illustrated in Figure 6.12 (from Granot, Ramirez-Ruiz, & Perna 2005), giving the emission diagram of a uniform relativistic jet, with sharp edges and half-opening angle $\theta_j$, seen by an off-axis observer whose line of sight makes an angle $\theta_{\text{obs}} > \theta_j$ with the jet axis. This schematic diagram explains why the light-curve rises and peaks when $\Gamma$ drops to $\sim (\theta_{\text{obs}} - \theta_j)^{-1}$ as the line of sight enters the beaming cone of the emitting material at the edge of the jet (the middle cone in the figure) and later decays with time.

### 6.7.2 GRB beaming and energetics

According to the previous section, the measure of $t_b$, the time of the jet break, is crucial for inferring $\theta_j$, the beaming angle of the GRB, and its energy output $E_\gamma$. As explained in Section 3.4, $t_b$ has a strong dependence on $\theta_j$: $t_b \propto \theta_j^{8/7}$ (Rhoads 1999, Panaitescu & Mészáros 1999, Sari, Piran, & Halpern 1999). $\theta_j$ is a decisive parameter because, along with the redshift, it is needed to calculate the total energy radiated by a GRB.

From the analyses in Section 6.7.1, it is, however, clear that the interpretation of the breaks in afterglow light-curves is still a problem with a lot of theoretical debate. As we have seen such breaks can have many causes beyond the jet geometry, like for instance variations of the external density, a break in the power-law distribution of radiating electrons or a transition from relativistic to non-relativistic regime (Kumar & Panaitescu 2000b, Li & Chevalier 2001, Wang, Dai, & Lu 2000b). In short, identification of a jet and determination of the beaming angle are difficult to achieve.
Figure 6.13. Histogram of beaming-corrected GRB energies, $E_\gamma$, with three equal logarithmic spacings per decade. The sample includes 24 GRBs: 9 GRBs with measured redshift, jet break time and the density of the surrounding medium, 7 GRBs with measured redshift and the jet break time, and 8 GRBs with measured redshift and a constraint on the jet break time (the arrows indicate upper or lower limits on $E_\gamma$ for these GRBs). The histogram shows a narrow distribution of GRB energies about the mean value $E_\gamma = 1.3 \times 10^{51}$ erg (from Bloom, Frail, & Kulkarni 2003).

In spite of these difficulties, a simple jet model where the half opening angle is determined using the relation $t_j \propto \theta_j^{8/3}$ given by Sari, Piran, and Halpern (1999), allowed Frail et al. (2001), to establish that the gamma-ray energy release $E_\gamma$ is narrowly clustered around $5 \times 10^{50}$ erg (Figure 3.11), assuming a uniform, or top-hat beam. Here $E_\gamma$ is the energy released in gamma-rays, corrected for the beaming, $E_\gamma = f_b E_{\text{iso}}(\gamma)$, with $f_b = (1 - \cos \theta_j) \sim \theta_j^2/2$. A more recent analysis by Bloom, Frail, and Kulkarni (2003) confirmed this clustering around $\sim 1.3 \times 10^{51}$ erg for 24 GRBs (Figure 6.13). Panaitescu and Kumar (2001a) and Panaitescu (2005a,b) performed the broadband modeling of 8 and 10 GRB afterglows respectively and found outflow energies ranging between 2 and $6 \times 10^{50}$ erg.

It is interesting to see that good estimates of the total energy of the relativistic ejecta, $E_{\text{rel}}$, emitted by the central engine have also been obtained with the observed broadband emission of GRB afterglows (Panaitescu & Kumar 2001a, Piran et al. 2001). These authors have shown that $E_{\text{rel}}$ can be estimated within a factor of 2 by the determination of $E_{\text{ka}}$, the kinetic energy of the ejecta during the adiabatic afterglow phase. In fact, $E_{\text{rel}} \geq E_\gamma + E_{\text{ka}} = \varepsilon E_{\text{rel}} + E_{\text{ka}}$. $E_\gamma$ is the total energy radiated in high-energy photons ($E_\gamma = (\theta^2/2)E_{\text{iso}}$, where $\theta$ is the effective angle of the $\gamma$-ray emission and $E_{\text{iso}}$ the isotropic equivalent energy emitted in $\gamma$-rays). The inequality arises from possible energy losses during the early afterglow radiative phase. These losses seem to be small (Giblin et al. 1999, Tkachenko et al. 2000). Therefore with $\varepsilon \sim 10\%$ we expect to have $E_{\text{ka}} \sim E_{\text{rel}}$ within a factor of 2. As noted by Piran et al. (2001), $\varepsilon$ cannot be too close to unity otherwise there would be no afterglow. These authors determined the spread of $E_{\text{ka}}$ using the X-ray afterglow fluxes. They found that $E_{\text{ka}}$ is quasi-constant. This is a good indication that the total energy emitted by the central engine does not vary significantly and that it is close to the energy estimated through
\( E_{\text{ka}} \). The energy of the GRB jets span less than one order of magnitude and the mean value given by Panaitescu & Kumar (2001a) is about \( 3 \times 10^{50} \) erg.

Piran et al. (2001), considering that this determination of \( E_{\text{ka}} \) is cumbersome and time-consuming, proposed to measure the width of the energy distribution using the X-ray afterglow flux at a fixed time after the explosion. They believed that the best choice was to use the X-ray afterglow flux between several hours and one day after the explosion. They showed that the distribution of the observed X-ray flux of GRB afterglows is narrow, implying that the energy of the GRB jets spans less than one order of magnitude after the early afterglow phase. This constancy of \( E_{\text{ka}} \), and the fact that it appears to be a rather good estimate of \( E_{\text{rel}} \) might indicate that the total energy emitted by the central engine is almost constant, with the reserve that we have no indication of the amount of energy which can be emitted in other forms: neutrinos, non-relativistic particles, gravitational waves, etc. This nearly constant energy reservoir might also indicate (Piran et al. 2001) that in the collapsar model of long GRBs (see Chapter 8), the mass accretion onto the BH and the conversion of the rotational energy of the BH to the kinetic energy of the jet, do not vary much from one burst to another. This seems at odds with the fact that both the disk mass and the BH spin are expected to vary widely in the collapse of massive stars.

Berger, Kulkarni, & Frail (2003) contested the statistical approach of Piran et al. (2001) (see the arguments in their paper). Using 41 GRB afterglows and their jet opening angle \( \theta_j \), they showed that there is a significant dispersion in the X-ray fluxes normalized at 10 h. However, as there is a strong correlation between the X-ray flux and the beaming fraction \( f_{\text{b}} \approx 1 - \cos(\theta_j) \), the true X-ray luminosity of GRB afterglows \( L_{\text{x}} \) must be approximately constant. The strong clustering of \( L_{\text{x}} \) implies that \( E_{\text{ka}} \), the adiabatic blastwave kinetic energy in the afterglow phase, is tightly clustered, as concluded by Piran et al. (2001). It seems therefore that there is a standard kinetic energy reservoir in GRBs and that the energy imparted by the GRB central engine to the relativistic ejecta is approximately constant. The authors concluded that the constancy of both \( E_\gamma \) and \( E_{\text{ka}} \) indicates that GRB jets must be relatively homogeneous and maintain simple conical geometry all the way from the internal shock (\( 10^{14} \)–\( 10^{15} \) cm) to the epoch of the jet break (\( 10^{17} \) cm). They also ruled out the patchy shell model proposed by Kumar & Piran (2000a) in which brighter bursts are due to bright spots along specific lines of sight.

The energies obtained through \( E_{\text{ka}} \) can be compared with the values of \( E_\gamma \) reported by Frail et al. (2001) and Bloom, Frail, and Kulkarni (2003). Using 17 GRBs with a redshift and a jet break in their optical afterglow, Lloyd-Ronning and Zhang (2004) showed that overall the radiated energy \( E_\gamma \) and the kinetic energy \( E_{\text{ka}} \) appear to be highly correlated.

Figure 6.14 (from Berger et al. 2003) is interesting in this context as it shows the distribution of \( E_\gamma, E_{\text{ka}}, L_{\text{x}} \) (the kinetic energy derived from the X-ray luminosity of the afterglow at 10 hours), and the total energy \( E_{\text{rel}} = E_{\text{ka}} + E_\gamma \) (where \( E_{\text{ka}} \) is computed from a complete modeling of the afterglow) for 18 GRBs. This figure shows that the total energy of the relativistic ejecta \( E_{\text{rel}} \) seems nearly constant, and that \( E_\gamma \) and \( E_{\text{ka}} \) provide poor estimates of \( E_{\text{rel}} \).
Figure 6.14. Distribution of various energy estimates for 18 GRBs (Berger et al. 2003). The upper panel shows $E_\gamma$ the distribution of the beaming corrected gamma-ray energy. The middle panel shows $E_{k,x}$ the kinetic energy inferred from X-rays at $t = 10$ h. The lower panel shows $E_{\text{rel}} = E_\gamma + E_{\text{k,rad}}$ the total relativistic energy, where $E_{\text{k,rad}}$ (also called $E_{\text{k,a}}$ in the text) is the beaming-corrected kinetic energy inferred from the broadband afterglow. There is a significantly wider dispersion in $E_\gamma$ and $E_{k,x}$ as compared to the total explosive yield $E_{\text{rel}}$. For Berger et al. (2003) this indicates that engines in cosmic explosions produce approximately the same quantity of energy. They also point out that the quality of GRB engines, as indicated by ultra-relativistic output ($E_\gamma$ and $E_{k,x}$) varies widely.

The correlation between the total energy of GRBs and their jet break angles has motivated the search for a similar relation between the peak luminosity $L$ and the jet angles $L \propto \theta^{-2}$. This relation implies that the GRB luminosity function simply reflects the distribution of beaming angles. Using the sample of Bloom, Frail, and Kulkarni (2003), Guetta, Piran, and Waxman (2005) have shown that the energy–angle relation found by Frail et al. (2001) also applies to the peak luminosity and that there is a clustering of the true peak luminosities around $\sim 3.2 \times 10^{49}$ erg s$^{-1}$. Taking into account the observed luminosity function and an assumption about the star-formation rate, this correlation allows derivation of the beaming angle, which agrees well with the one derived by Bloom, Frail, and Kulkarni (2003). Guetta, Piran, and Waxman (2005) recalculated the average beaming factor $f_b^{-1} \sim (\theta_j^2/2)^{-1}$, finding a value of $\langle f_b^{-1} \rangle = 75 \pm 25$. This beaming factor gives, for every observed GRB, the average number of bursts that are missed because their jet is not directed towards the Earth. Frail et al. (2001) estimated this factor to be $\sim 500$, while Van Putten and Regimbau (2003) gave $f_b^{-1} \sim 450$. Guetta, Piran, and Waxman (2005) proposed an explanation for their much lower value: they used a weighted average of the predicted
angular distribution while Frail et al. (2001) and Van Putten and Regimbau (2003) used a simple average over the ‘true’ angular distribution.

Even though there is a significant spread in the results, these analyses suggest that the energy output of the central engine could be a universal value for long-duration GRBs and that the typical opening angles for the collimated bursts are in the range 1° to 20° (Piran et al. 2001).

6.7.3 Orphan afterglows

Since the vast majority of GRB jets are not directed towards the Earth, most GRBs are seen off-axis. Observers who are not too far off-axis can detect the late afterglows of these ‘invisible’ GRBs, which have been called ‘orphan afterglows’ (see Figure 6.12 for an illustration). The GRB collimation can thus be constrained by searches of orphan afterglows at X-ray, visible or radio wavelengths.

A search for orphan afterglow candidates at X-ray wavelengths was performed by Greiner et al. (2000) in the Rosat all-sky survey data. This search revealed no convincing candidate, leading the authors to conclude that the beaming angle is comparable in the X-ray and gamma-ray domains. As the Rosat all-sky survey was not very sensitive, it could only permit the detection of bright afterglows, in the first few hours after the GRB when the jet has not spread out significantly. The result of Greiner et al. (2000) appears compatible with this picture.

At radio wavelengths, Gal Yam et al. (2006) presented radio and optical follow-up observations of nine possible radio transients discovered by Levinson et al. (2002) in a survey covering ∼1/17 of the sky. They place an upper limit of 65 radio transients per year above 6 mJy over the entire sky. One implication of this search is a limit on the typical beaming factor of the GRBs: $f_b^{-1} \geq 60$.\(^2\)

Constraints on visible orphan afterglows have been obtained with ROTSE III (Rykoff et al. 2005), with the 2.2-m MPI/ESO telescopes (Rau, Greiner, & Schwartz 2006), and with the CFHTLS very wide survey (Malacrinò et al. 2006). None of these searches has found any serious candidate. The constraints from the different surveys are shown in Figure 6.15 (Malacrinò et al. 2007) by the maximum number of afterglows allowed over the entire sky versus the magnitude. The comparison is made with the predictions of various models (Totani & Panaitescu 2002, Nakar, Piran, & Granot 2002, Zou, Wu, & Dai 2007). The non-detection of orphan afterglows in the CFHTLS Very Wide Survey implies that at magnitude $r = 22.5$, orphan afterglows can be at most 13 times more frequent than afterglows of on-axis GRBs (Malacrinò et al. 2007). The comparison of these constraints with the predictions of the models (Figure 6.15) shows that the current searches are too shallow to significantly constrain the models.

\(^2\)Interestingly, radio observations provide a lower limit on the beaming factor. Radio afterglows are observed in the non-relativistic regime, when they radiate isotropically. Hence, radio observations constrain the total number of GRBs within the volume accessible to the observations (which are flux-limited). Radio observations thus provide an upper limit on the space density of GRBs or a lower limit on the beaming factor.
Figure 6.15. Constraints on the rate of visible afterglows from GRBs obtained with the CFHTLS Very Wide Survey. The large triangle indicates that there cannot be more than 220 visible GRB afterglows brighter than magnitude 22.5 at a given time in the all sky. Other constraints, obtained with ROTSE-III (Rykoff et al. 2005) the MPI/ESO survey (Rau, Greiner, and Schwarz 2006) and the DLS survey (Becker et al. 2004) are also shown. The upper limits on the frequency of orphan afterglows appear compatible with the current theoretical predictions, which are shown with lines (from Malacrino et al. 2007).

In spite of the success of the studies that support the presence of a jet, the real jet structure is not established (see also the situation of jet breaks after the first results of Swift in Chapter 7), so it is not unreasonable to imagine jet structures that differ from the simple uniform jet.

6.7.4 The structure of the jet

Although the uniform jet (UJ) model is the simplest one that can describe the structure of the jet, other models have been introduced, and among them the universal structured jet model (USJ; Lipunov, Postnov, & Prokhorov 2001, Rossi, Lazzati, & Rees 2002, Zhang & Mészáros 2002a). In the uniform jet model different values of the jet break time \( t_b \) are observed, mainly due to different beaming angles \( \theta_j \). In the USJ model all GRB jets have identical angular profiles and total energy (they have a universal angular distribution of jet energy). In this model, the different values of \( t_b \) arise mainly due to different viewing angles \( \theta_{\text{obs}} \) from the jet axis. The light-curve is similar to that of a uniform jet with an opening angle \( \theta_j = \theta_{\text{obs}} \), and an energy per
unit solid angle equal to the energy per solid angle of the USJ at \( \theta = \theta_{\text{obs}} \). The USJ structure for all GRBs is described by a function \( \varepsilon(\theta) \) outside of some core angle \( \theta_c \), \( \varepsilon = \varepsilon_c (\theta/\theta_c)^{-k} \) for \( \theta > \theta_c \). In this type of jet the kinetic energy per unit solid angle \( \varepsilon \) and the Lorentz factor \( \Gamma \) vary smoothly, and Mészáros, Rees, and Wijers (1998) proposed that \( \varepsilon \) and \( \Gamma \) vary as power-laws as a function of the angle \( \theta \) from the jet axis. Rossi, Lazzati, and Rees (2002) proposed the following power-law dependence: \( \varepsilon = \varepsilon_c \) for \( \theta < \theta_c \), and \( \varepsilon \propto \theta^{-2} \) for \( \theta_c < \theta < \theta_j \), with a power-law dependence also for \( \Gamma \). With this law all GRBs have the same total energy \( E_j = 2\pi\theta_j^2 \varepsilon = \text{const.} \) These configurations are characterized by a strong anisotropy of the angular distribution of the fireball energy around the jet axis. This axis is physically related to the rotational axis of the central engine (see the collapsar model, Chapter 8), so it is reasonable to assume that a higher energy concentration is found close to the jet axis (Zhang & Mészáros 2002a). At the center of the jet, the ejecta will be faster and lighter, with higher Lorentz factor \( \Gamma \). At larger angles from the jet core the ejecta will interact with the stellar material (`a wall`) and thus will be slower and will entrain more baryons (lower \( \Gamma \) values are expected). This universal jet model may appear to be a more natural representation of what can be expected in nature. Lipunov, Postnov, and Prokhorov (2001), considered another type of `universal' jet, with a three-step profile: a spherical shell, a 20° cone and a 3° cone.

In fact both uniform and universal jet structures can explain the observed correlation between \( E_{\text{iso}} \) (the isotropic equivalent energy released in the prompt phase) and the jet break time \( t_b \) in the optical afterglow light-curve: \( t_b \propto E_{\text{iso}}^{-1} \) (Frail et al. 2001, Bloom, Frail, & Kulkarni 2003). In the USJ model this correlation can be reproduced if the jet structure follows \( \varepsilon \propto \theta^{-2} \) (Rossi, Lazzati, & Rees 2002, Zhang & Mészáros 2002a, Salmonson 2003). In the UJ model \( t_b \) depends mainly on \( \theta_j \) (Rhoads 1997) while in the USJ model it depends mainly on the viewing angle \( \theta_{\text{obs}} \); moreover this model predicts a more gradual variation of the break than the uniform jet model. Gradual and smooth breaks seem to be consistent with the observations of several well-studied GRB optical afterglows (Zhang & Mészáros, 2002a).

Concerning the USJ model, Granot and Kumar (2003) have also shown that the jet structure with \( \varepsilon \propto \theta^{-2} \) is required to reproduce typical afterglow light-curves. However, like Rossi, Lazzati, and Rees (2002), they observe that more afterglow observations are needed to conclude definitely on the geometry of the jet. Using a sample of 28 bursts with \( \theta_{\text{break}} \) from Bloom, Frail, and Kulkarni (2003), Perna, Sari, and Frail (2003) used the observed GRB intensity distribution to infer the distribution of observing angles for the structured jet model. They predicted the most common angle to be \( \sim 0.12 \text{ rad} \) (\( \sim 7^\circ \)) and they find a rough agreement with the sample. We can also add that after examining several afterglows with breaks Panaitescu & Kumar (2003) found that outflows endowed with structure are required only if the circum-burst medium has a wind-like profile. They had previously noted that most well-observed afterglows are well fitted by a uniform jet interacting with a homogeneous external medium (Panaitescu & Kumar 2002; but see Zeh, Klose, & Kann 2006).

Less positive conclusions concerning the USJ model are given by Nakar, Granot, and Guetta (2004). These authors showed that a sample of 16 GRBs with known
angle and redshift strongly disfavors the USJ model, but since the sample is very inhomogeneous (involving many different instruments) it cannot be used to definitely rule out the USJ model. Hence in 2004, before the launch of Swift, the structure of GRB jets remained an open question.

These uncertainties over the structure of the jets may explain why other models less popular that the UJ and USJ models have been proposed. Among them

- A jet with a Gaussian angular profile (Kumar & Granot 2003, Zhang & Mészáros 2002a), which may represent a more realistic and more physical version of the uniform jet, with smooth rather than sharp edges. A Gaussian function can be considered as approximately intermediate between the UJ and USJ models but closer to the UJ model since the energy in the wings is much smaller than in the core. Quasi-universal Gaussian jets with typical opening angle of $5.7^\circ \pm 3.4^\circ$ degrees and a standard jet energy of about $\log E$ (erg) = 51.1 ± 0.3 are compatible with the observations of both GRBs and XRFs (Zhang et al. 2004). In agreement with Lloyd-Ronning, Dai, and Zhang (2004) they find that both Gaussian and power-law jet structures can reproduce the data with only minimal scatter in the model parameters.

- A two-component jet (Pedersen et al. 1998, Frail et al. 2000a, Berger et al. 2003, Huang et al. 2004, Peng, Königl, & Granot 2005, Wu et al. 2005a). This model includes a narrow uniform jet of initial Lorentz factor $\Gamma_0 > 100$ surrounded by a wider uniform jet with $\Gamma_0 \sim 10$–30. Motivations for such a structure have been found in the context of the cocoon in the collapsar model (Ramirez-Ruiz, Celotti, & Rees 2002) as well as in the context of a hydro-magnetically driven jet originating from a neutron star or from a neutron-rich accretion disk formed in the collapse of a massive star (Vlahakis, Peng, & Königl 2003). The light-curves have been calculated analytically (Peng, Königl, & Granot, 2005) or semi-analytically (Huang et al. 2004, Wu et al. 2005a). Such a model might account for the fast rebrightening episode seen in the optical afterglow of GRB 030329 at 1.5 days, which could be due to the deceleration of the slower jet component (Figure 6.16; Berger et al. 2003). But this conclusion is in contradiction with the study of two-component jet models by Granot (2005). Using a simple formalism Granot (2005) concluded that the two-component model cannot produce sharp features in the afterglow and that it cannot explain the sharp rebrightening observed in the optical afterglow of GRB 030329. Nevertheless, among other results, the millimeter observations of GRB 030329 support the two-component jet-like outflow proposed by Berger et al. (2003) in which a narrow jet, $5^\circ$, is responsible for the high-energy emission and the early optical afterglow, while a broad jet, $17^\circ$, carrying most of the energy, powers the radio and late optical afterglow emission (after 1.5 days). The millimeter observations are not compatible with the refreshed shock model (Sheth et al. 2003). Without being exhaustive, we can add that numerical simulations by Zhang, Woosley, and McFadyen (2003) indicate that a double Gaussian structure for the jet may provide the best description when one Gaussian is used for the core of the jet and one for the wings.
Figure 6.16. Radio to X-ray light-curves of the afterglow of GRB 030329 (from Berger et al. 2003). The authors propose to interpret these observations with a two-component jet. In this interpretation the sharp increase in the optical flux at $t \sim 1.5$ days is due to the deceleration of the slower second jet component. The dot dashed line is the optical emission from SN 1998bw at the redshift of GRB030329 ($z \sim 0.168$). The solid line fitting R band data is the combination of the two jet components and the SN. In the upper right of the figure the radio data are also shown with a break at $t_{j,\text{rad}} \sim 10$ days; radio emission is dominated by the wide and slow jet component. The X-ray afterglow is shown as gray stars; at these energies the emission is dominated by the narrow jet component.

- More ‘exotic’ jet structures have been proposed. One example is a jet with a ring-shaped cross-section, referred to as a ‘hollow cone’, that is uniform within $\theta_e < \theta < (\theta_e + \Delta \theta)$ with $\Delta \theta \ll \theta_e$ (Eichler & Levinson 2004, Levinson & Eichler 2004, Lazzati & Begelman 2005). Another example is a ‘fan’ or ‘sheet’-shaped jet (Thompson 2005) where a wind from a proto-neutron star becomes relativistic as the density in its immediate vicinity drops and might form a thin sheet of relativistic outflow that is able to penetrate through the progenitor star along the rotational equator to form a relativistic outflow within $\Delta \theta \ll 1$ around $\theta = \pi/2$ (or $\theta_e = \pi/2 - \Delta \theta/2$).

In a paper devoted to the calculation of afterglow light-curves Granot (2005) gives a schematic diagram of the energy per solid angle $dE/d\Omega$ versus $\theta$ for these various jet structures (his Figure 1) and predicts the evolution of their optical afterglow (his Figures 2 to 8). An exhaustive analysis of our understanding of this crucial
issue of jets and their models has been given by Granot (2007). To summarize his conclusions a very important point seems to be the modest degree of lateral expansion of the jet as long as it is relativistic, the full hydrodynamics simulations which have been used being considered as the most reliable. This result is reported for uniform jets as well or for Gaussian or structured jets. Different approaches have been used to constrain these jet models (see Granot 2007) without reaching definite conclusions. Even if the shape of afterglow light-curves is in principle a relatively robust diagnostic tool for constraining the jet structure, various authors have shown that it is in fact extremely difficult to infer the structure of the jet based on the observed light-curves only (Rossi et al. 2002, Granot & Kumar 2003, Kumar & Granot 2003, Salmonson 2003, Rossi et al. 2004). As we report in the next section, polarization data can also be used to discriminate the models (see, for instance, Lazzati et al. 2004 for the case of GRB 020813). But the problem is also to constrain the magnetic field configuration in the emitting region and it appears difficult to constrain each of these ingredients separately. This means that strong assumptions have to be made on the magnetic field configuration in order to infer the jet structure, and vice versa. So let us go to the polarization observations and discuss the additional information they provide.

6.7.5 Polarization: a tool to distinguish the different jet models

Synchrotron emission from a jet in which the spherical symmetry is broken would naturally produce polarized emission (Ghisellini & Lazzati 1999, Gruzinov 1999, Sari 1999). The level and the direction of the polarization are expected to vary with time and to give indications on the geometry of the emitting jet. Rossi et al. (2004), for instance, have shown that polarization curves are extremely different for different types of jets. Their Figure 18 shows a comparison of the light-curves and polarization curves for a universal structured jet (USJ), for a Gaussian jet (GJ), and for a uniform jet (UJ). So, what polarization can robustly determine is whether the energy distribution in the jet is uniform or centrally concentrated. Unfortunately, the authors stress that, due to many uncertainties inherent in the derivation of polarization curves, it is hardly possible to use them to accurately measure the energy distribution of the jet. In addition, Rossi et al. (2004) showed that the degree of polarization is quite variable with time and is particularly significant around the time of the jet break. The position angles at that time depend strongly on the geometry of the jet.

Polarization measurements are also important because they are directly connected with the issue of magnetic fields in GRBs, which remain a truly unsolved question. The origin and the geometry of the magnetic field, and its coherence properties are open questions, and various theoretical approaches to describe the magnetic field configuration have been proposed. Among them Gruzinov and Waxman (1999) and Gruzinov (1999) estimated a polarization around 10%, based on the assumption that the overall emission reaching the observer is produced in a finite number of causally disconnected regions \( N \sim 50 \), each of which is embedded in a completely ordered magnetic field. Another magnetic field generation process was proposed by Medvedev and Loeb (1999). They consider a magnetic field completely tangled in the plane of the shock front but with a high degree of coherence in
the orthogonal direction. If the fireball is spherical no polarization is expected, except if part of the fireball emission is amplified and part of it is obscured producing an asymmetry. In these models the required degree of asymmetry is in the magnetic field configuration and not in the overall geometry of the fireball. On the other hand, Granot and Königl (2003) proposed that the required asymmetry might be provided by an ordered magnetic field embedded in the circum-burst material. This ordered magnetic field could be amplified when the shock propagates in it, to reach values close to equipartition with the energy density of the shocked material, like in a ‘pulsar wind bubble’ scenario (Königl & Granot 2002). Ghisellini and Lazzati (1999) have considered an alternative model with a magnetic field geometry analogous to that of Medvedev and Loeb (1999) but in a fireball which is collimated in a cone and an observer which is slightly off-axis. The circular symmetry is broken and a net polarization can result. The polarization curves present two maxima (their Figure 4) and the polarization angle is predicted to change by 90° between the first and the second maximum. Other models predicting significant polarization have been proposed, see for instance Levinson and Eichler (2004).

Let us see now how these predictions compare with the observational results. Early attempts to measure optical polarization gave rise to an upper limit of 2.3% for GRB 990123 (Hjorth et al. 1999). Upper limits were also obtained for radio polarization measurements: 19% for GRB 980329 (Taylor et al. 1998), and 8% for GRB 980703 (Frail et al. 1998). The discovery of a significant, albeit small (1.7 ± 0.2% measured 0.77 days after the burst), degree of linear polarization was reported by Covino et al. (1999a) in the optical afterglow of GRB 990510. This observation was confirmed by Wijers et al. (1999), who obtained 1.6 ± 0.2%, 0.86 days after the trigger. After 1.81 days, the level of polarization was still consistent with the same value but with much larger uncertainties. Several detections of linear polarization in a number of afterglows have followed, in particular in GRB 990712 (Rol et al. 2000, Björnsson et al. 2000), GRB 020405 (Bersier et al. 2003a, Covino et al. 2003), GRB 020813, and GRB 030329 (see for example Covino et al. 2005). The polarized flux is always small, \( P < 3\% \), with a constant or smoothly variable level. For GRB 020813 the polarization was observed over a significant duration (Covino et al. 2005, Lazzati et al. 2004), and for the first time the polarization was measured before and after the break in the optical light-curve (Gorosabel et al. 2004). There is no large variation of the polarization, which changes from \(~2.7\%\) to \(1.3\%\) (Figure 6.17, from Lazzati et al. 2004; see also Figure 2 of Gorosabel et al. 2004). The same weak variation is also valid for the polarization angle.

For these bursts it has been possible to rule out, in most cases, that the observed polarization was induced by dust along the line of sight in our own Galaxy. Another possible source of polarization could, however, be intervening dust in the host galaxy. The pollution by these two sources of polarization has been discussed by Lazzati et al. (2003) for GRB 021004. They stressed the need for good quality spectro-polarimetry and multi-time/multi-filter imaging polarimetry, especially the need for late-time measurements. These are the conditions for disentangling the polarization of the GRB afterglow from the polarization due to dust in our Galaxy and in the host galaxy.
Figure 6.17. Polarization of the optical afterglow of GRB 020813. Part (a) shows the observed polarization (level and position angle) around the jet break time $t_b$ with the predictions of various jet models. Part (b) shows the evolution of the models over a wider range of times, particularly at early times. Models yielding an acceptable fit are plotted with a thick line. The gray-shaded area shows the acceptable range for the jet break time. The models are labelled by two letters: the first characterizes the nature of the jet (Hydrodynamic or Magnetized) and the second its geometry (Homogeneous or Structured). HHJ is the standard top-hat jet with the energy uniformly distributed within the jet. HSJ, the hydrodynamic structured jet has more energy in the centre and less in the wings, this model is usually called USJ. MHJ, the magnetic homogeneous jet is a jet in which the magnetic field has a toroidal structure: the energy is uniform within the jet as for the HHJ model. MSJ is the magnetized structured jet, the magnetic field is toroidal as in the case of MHJ but the energy distribution is inhomogeneous. The difference between the models are significant but mainly at early times when $t < t_b$ (from Lazzati et al. 2004).
Concerning GRB 021004, Rol et al. (2003) showed a 45° change in the polarization angle from 9 h after the burst to 16 h, and up to 90° using data at a later epoch. The degree of polarization shows no significant change. This result was used to rule out the structured jet model for this burst, because this model does not predict a change in the polarization angle. Other polarization measurements showing large variations in the degree of linear polarization within short timescales (Bersier et al. 2003a, Covino et al. 2004) cannot be reconciled with any of the current models.

To summarize the measurements of polarization made in the optical band it can be said that when the optical afterglow has a relatively smooth light-curve the polarization is observed to be low, at a level of 1–3%, with constant or smoothly variable level and position angle. When the light-curve exhibits deviations from smooth power-law decay, with bumps or rebrightenings, the polarization curves show a certain degree of complexity (e.g. GRB 021004; Lazzati et al. 2003, Rol et al. 2003), and significant variability (GRB 030329; Greiner et al. 2003). Covino et al. (2004, 2005) have given a complete review of the observations through the end of 2002. These observations are summarized in Figure 6.18 and in Table 6.4 (the Table 1 of Covino et al. 2004 which gives all the references).

Until now we have considered the polarization of the optical GRB afterglow, with relatively low levels reaching a few percent. This is much less than the surprisingly large value $p = 80\% \pm 20\%$ reported by Coburn and Boggs (2003) for the prompt γ-ray emission of the extraordinary bright GRB 021206, observed with the gamma-ray detectors of RHESSI. This large polarization could be explained (up to 60%) by an ordered transverse magnetic field advected from the central source (Granot & Königl 2003). Lyutikov, Pariev, and Blandford (2003) claimed that this high polarization favors Poynting-dominated outflows in GRBs. Waxman (2003) proposed another scenario, which does not require an ordered magnetic field. He assumed that the GRB outflow is a uniform jet with sharp edges with an opening angle $\theta_j \lesssim 1/\Gamma$, where $\Gamma \gtrsim 100$ is the Lorentz factor during the GRB. Our line of sight is assumed to be slightly outside the edge of the jet: $\theta_j < \theta_{\text{obs}} < (\theta_j + 1/\Gamma)$. In such a configuration the observer is able to see the bright GRB and a large polarization (see also Granot et al. 2002). These scenarios (ordered magnetic field vs ad hoc line of sight) have been compared by Granot (2003). He concluded that the most natural way to produce such a high polarization during the prompt GRB phase is to assume an ordered magnetic field carried out with the ejecta from the central source. A magnetic field generated at the shock itself seems less likely. But if the true polarization is smaller, about 20–30%, it could be explained by a magnetic field generated in internal shocks if one assumes that the magnetic field produced in internal shocks is significantly more anisotropic than in the afterglow (Granot 2003).

It should be noted, however, that the initial claim for a nearly 100% polarization in GRB 021206 was disputed by various authors. Careful reanalyses of the same data produced results that were much less significant. Wigger et al. (2004) showed that the level of polarization could have any value between 0% and 100% and concluded that the quality of the data was insufficient to constrain the true degree of polarization of GRB 021206. Rutledge and Fox (2004) reached a similar conclusion after they computed the number of photons available for the measurement of polarization,
Table 6.4. The results of the 25 polarization measurements performed on 9 GRBs before January 2003. 1 σ errors and 95% confidence level upper limits are reported. SP in column 5 stands for spectropolarimetry. For GRB 020813 and GRB 021004 the reported polarization degrees and position angles are not corrected for Milky Way interstellar matter–induced polarization (from Covino et al. 2004).

<table>
<thead>
<tr>
<th>Burst</th>
<th>P  (%)</th>
<th>δ  (°)</th>
<th>Δt (h)</th>
<th>Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRB 990123</td>
<td>&lt;2.3</td>
<td></td>
<td>18</td>
<td>R</td>
</tr>
<tr>
<td>GRB 990510</td>
<td>1.7 ± 0.2</td>
<td>101 ± 3</td>
<td>18</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>1.6 ± 0.2</td>
<td>96 ± 4</td>
<td>21</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>&lt;3.9</td>
<td></td>
<td>43</td>
<td>R</td>
</tr>
<tr>
<td>GRB 990712</td>
<td>2.9 ± 0.4</td>
<td>122 ± 4</td>
<td>11</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>1.2 ± 0.4</td>
<td>116 ± 10</td>
<td>17</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>2.2 ± 0.7</td>
<td>139 ± 10</td>
<td>35</td>
<td>R</td>
</tr>
<tr>
<td>GRB 991216</td>
<td>&lt;2.7</td>
<td></td>
<td>35</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>&lt;5</td>
<td></td>
<td>60</td>
<td>V</td>
</tr>
<tr>
<td>GRB 010222</td>
<td>1.4 ± 0.6</td>
<td></td>
<td>22</td>
<td>V</td>
</tr>
<tr>
<td>GRB 011211</td>
<td>&lt;2.0</td>
<td></td>
<td>37</td>
<td>R</td>
</tr>
<tr>
<td>GRB 020405</td>
<td>1.5 ± 0.4</td>
<td>172 ± 8</td>
<td>29</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>9.8 ± 1.3</td>
<td>180 ± 4</td>
<td>31</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>2.0 ± 0.3</td>
<td>154 ± 5</td>
<td>52</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>1.5 ± 0.4</td>
<td>168 ± 9</td>
<td>78</td>
<td>V</td>
</tr>
<tr>
<td>GRB 020813</td>
<td>2.3 – 3.1</td>
<td>153 – 162</td>
<td>6</td>
<td>SP</td>
</tr>
<tr>
<td></td>
<td>1.2 ± 0.2</td>
<td>158 ± 5</td>
<td>24</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>1.6 ± 0.3</td>
<td>163 ± 6</td>
<td>29</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>2.0 ± 0.4</td>
<td>179 ± 6</td>
<td>50</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>4.3 ± 1.7</td>
<td>177 ± 11</td>
<td>96</td>
<td>V</td>
</tr>
<tr>
<td>GRB 021004</td>
<td>&lt;5</td>
<td></td>
<td>11</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td>1.3 ± 0.1</td>
<td>114 ± 2</td>
<td>15</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>1.3 ± 0.3</td>
<td>125 ± 1</td>
<td>16</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>1.4 – 2.3</td>
<td>111 – 126</td>
<td>19</td>
<td>SP</td>
</tr>
<tr>
<td></td>
<td>0.7 ± 0.2</td>
<td>89 ± 10</td>
<td>91</td>
<td>V</td>
</tr>
</tbody>
</table>

which appeared to be too low to draw any conclusion. The polarization of the prompt gamma-ray signal from the very bright GRB 041219A was also studied with the germanium spectrometer of INTEGRAL (McGlynn et al. 2007, Kalemci et al. 2007). Both authors found a polarization level in the range 50% to 100% during the brightest part of the burst, but they concluded that the statistical significance of the signal was too low to draw firm conclusions, and that there may be unknown
systematics which could mimic GRB polarization. Similar results were obtained by Willis et al. (2005) for two GRBs whose prompt emission was observed by BATSE directly and reflected off the Earth’s atmosphere, allowing the measurement of their polarization. More accurate measurements of the level of polarization of the prompt gamma-ray emission are thus badly needed. The detection of a high level of linear polarization in a bright GRB would be a very strong indication that the synchrotron emission is the dominant radiation mechanism in the prompt GRB emission.

6.8 INHOMOGENEOUS EXTERNAL MEDIUM: THE CASE OF A WIND

The evolution of a relativistic blastwave in a variable-density medium was studied by Blandford and McKee in 1976. The resulting synchrotron light-curve was first studied by Mészáros, Rees, and Wijers in 1998. These authors considered the case of a spherical fireball of energy $E$ and bulk Lorentz factor $\Gamma$ expanding into an external medium of density $n(R) \propto R^{-d}$, where $R$ is the distance from the center of the explosion. The shocked gas evolves according to the conservation law:

$$n \Gamma^{(1+a)} R^3 \propto \Gamma^{(1+a)} R^{(3-d)} = \text{cst}$$

where $a = 1(0)$ corresponds to the conservation of energy (momentum), i.e. to the adiabatic (radiative) regime. These regimes have to be understood in a global sense, applying to the entire shell, baryons, magnetic fields, electrons, etc. Since in the observer frame the time $t$ must satisfy the classical relation: $R \propto c t \Gamma^2$, we have $\Gamma \propto R^{-(3-d)/(1+a)} \propto t^{-(3-d)/(1+a-2d)}$.

In the case of a homogeneous medium ($d = 0$), we find again: $\Gamma \propto R^{-3}$ in the strong coupling radiative regime ($a = 0$), and $\Gamma \propto R^{-3/2}$ in the adiabatic regime ($a = 1$). The power law of the previous equation is valid as long as the shell is relativistic. The authors have considered the light-curves representative of spherical inhomogeneous and non-spherical inhomogeneous models. They cover a wide range of behaviors, which can be found in the diversity of the observed afterglows. Among the various circum-burst media, the case of a wind with $n(R) \propto R^{-2}$ is particularly interesting because long GRBs may be associated with the collapse of very massive stars (see Chapter 8). Chevalier and Li (1999, 2000) have studied this case in
detail. They have computed the evolution of $\nu_m$ and $\nu_c$ with time, and their dependence on the microphysical parameters: $\nu_m \propto (1 + z)^{1/2} \epsilon_B^{1/2} \epsilon_r^{1/2} E^{1/2} t^{-3/2}$ and $\nu_c \propto (1 + z)^{-3/2} \epsilon_B^{3/2} E^{1/2} \epsilon_r^{-1} A^{-2} t^{1/2}$, where $A$ is a parameter describing the strength of the wind emitted by the progenitor in the late phases of its evolution (Chevalier & Li 1999). Unlike the ISM case where $\nu_c \propto t^{-1/2}$, in the wind model $\nu_c$ increases with time, and the critical time for the fast to slow cooling transition is $t_{\text{mc}} \propto (1 + z) \epsilon_r \epsilon_B A$. The deceleration time, $t_{\text{dec}}$, is typically the duration of the burst, since for typical parameters the wind model corresponds to the so-called thick shell case (Chevalier & Li 2000, Kobayashi & Zhang 2003a; see also Zhang et al. 2006).

The closure relations for wind models turn out to be:

$$\alpha = \frac{1 - \beta}{2} \quad \text{for} \quad \nu_c < \nu < \nu_m \quad (\alpha = 3\beta/2 \quad \text{for homogeneous medium})$$

$$\alpha = \frac{3p - 2}{4}, \quad \beta = \frac{p}{2}, \quad \alpha = \frac{3\beta - 1}{2} \quad \text{for} \quad \nu > \max(\nu_m \nu_c)$$

(as in a homogeneous medium).

Table 6.3 (from Zhang et al. 2006) summarizes the temporal index $\alpha$ and the spectral index $\beta$ for various afterglow models.

Chevalier and Li (1999, 2000) showed that high initial densities in a wind circum-burst medium imply that the cooling frequency is initially low and it increases with time. Their main conclusions are the following:

- For typical parameters the fast-cooling phase lasts for 1–2 days, considerably longer than in the homogeneous interstellar case.
- The transition to non-relativistic evolution occurs at $\sim 2$ yr for typical parameters. This is much longer than in the case of an interaction with the ISM because of the low density of the outer parts of the wind.
- The optical synchrotron emission from the early reverse shock has a peak magnitude $\sim 12$ for typical parameters.
- The interaction with a wind cannot produce the temporal behavior of the observed optical flash in the case of GRB 990123. The case of interstellar interaction appears strong for this burst.
- There is strong evidence for wind interaction for the two famous bursts GRB 970228 and GRB 970508 (but Frail, Waxman, & Kulkarni (2000) rejected a wind model for GRB 970508). On the other hand, observations of the afterglows of GRB 990123 and GRB 990510 favor the interaction of the jet with the interstellar medium.
- The main result outlined by the authors is the fact that we have pieces of evidence for two types of environnement for GRB afterglows: a constant density interstellar medium or a wind typical of Wolf–Rayet stars. Using the fact that very early the pre-shock density is higher for the wind case, the authors predict a prompt optical emission that dies off faster in the wind case than in the interstellar case.

Kobayashi and Zhang (2003a) studied the prompt optical emission from the reverse shock in a wind environment. They also report the distinct evolutions of the reverse shock in wind and ISM models. Another way to distinguish between wind and ISM models is proposed by Panaitescu and Kumar (2000): they found that
observations at submillimeter frequencies during the first day would provide the best way to discriminate between the two models. They studied the light-curves produced when a jet propagates into a medium with a power-law density profile. They found that in the wind model little steepening of the light-curve is expected when $\Gamma \theta_j \sim 1$, even for highly collimated ejecta. This might explain the absence of jet breaks in many GRBs. The light-curves of the afterglow in the adiabatic, synchrotron-dominated, cooling regime have also been studied by Sari, Piran, and Narayan (1998) and Granot and Sari (2002). Wu et al. (2005b) computed the light-curves for a semi-radiative blastwave in a stellar wind, and emphasized the crucial role of inverse Compton cooling in this regime (see their Figures 1 and 2).

Even though the case of a wind in the circum-burst medium is very interesting and astrophysically justified, many other situations may exist. One of them is the case of GRBs going off in a very low-density environment, which can suppress or dilute their afterglows. Such GRBs, called ‘naked GRBs’ have been studied by Kumar and Panaitescu (2000a).

On the other hand, the case of a dense circum-burst medium with density $n \sim 10^7$–$10^8$ cm$^{-3}$ can also be considered (Dai & Lu 1999, 2000). In this case the shock wave rapidly evolves into the sub-relativistic phase after a very short relativistic phase. A break or a long-term steep decay are expected, depending on the time at which the shock starts entering into the subrelativistic phase. These possibilities were considered by Dai and Lu (2000) for two GRB 990123 and GRB 980519. The radio afterglow of GRB 980519 was also interpreted by the presence of such a dense medium (Wang, Dai, & Lu 2000b).

Finally, the case of an inhomogeneous circum-burst medium is also quite interesting as we explain in Section 6.8.1.

### 6.8.1 Inhomogeneities in the circum-burst medium

Afterglows with variable light-curves might be explained by ejecta encountering matter with an irregular density profile. Such a situation has been considered by Lazzati et al. (2002), Heyl and Perna (2003), and Nakar, Piran, and Granot (2003) after the discovery of GRB 021004 by HETE-2. The light-curves resulting from this complex situation can be approximated by considering a series of emissions using instantaneous Blandford–McKee solutions, each with its own external density. Such calculations were developed by Nakar and Piran (2003), and the formalism they have established can be used to calculate complete light-curves. For instance, they computed the light-curve resulting from a Gaussian overdensity ($\Delta R/R = 0.1$, with various levels of overdensity, and $\nu_m < \nu < \nu_c$) in the ISM, assumed to be typical of a clumpy environment (Figure 6.19). The result does not take into account the reverse shock resulting from density enhancement and its effect on the blastwave. In fact the authors showed that the observed flux depends on the external density $n$ roughly as $n^{1/2}$. Thus a large density contrast is necessary to produce a significant rebrightening. But such a contrast produces a strong reverse shock that will decrease the Lorentz factor $\Gamma_{sh}$ of the emitting matter behind the shock. The result is a sharp drop of the emission below $\nu_c$ and a long delay in the arrival time of the photons (the
observer’s time is $\sim \Gamma_{sh}^{-2}$). These two effects contribute to suppress the flux and to set a strong limit on the steepness of rebrightening events due to density variations (Piran 2005). Finally, it should be noted that the geometry of the interaction and of the radiation implies that inhomogeneities in the medium surrounding the burst cannot produce sharp features with typical timescales $\Delta T/T \ll 1$ before the jet break, even for density variations with $\Delta R/R \ll 1$.

Dai and Lu (2002) considered a circum-burst medium presenting jumps in density. In their model these authors considered only one density jump, a spherical relativistic blastwave, and they neglected the emission off the line of sight which can significantly reduce the initial abrupt drop of the flux density. They noted that such density jumps could account for humps observed in the light-curves of the optical afterglows of GRB 970508 or GRB 000301C, even if other interpretations are possible, such as refreshed shocks or microlensing events. Dai and Wu (2003) considered such density jumps to explain the unusual temporal feature of the optical afterglow of GRB 030226 which faded as $t^{-1/2} \sim 0.2$ days after the burst, rebrightened during the period $0.2-0.5$ days, and then declined as $t^{-2}$.

Lazzati et al. (2002) presented a model where the bumps in the afterglow light-curve of GRB 021004 were due to the interaction of the fireball with density...
enhancements (clumps of matter) in the ambient medium. The presence of shells surrounding the massive progenitor star has also been invoked to explain the absorption lines seen in the afterglow of this burst, the first case for which the optical spectrum is dominated by absorption lines from high-ionization species with multiple velocity components (Schaefer et al. 2003). For this same burst, Li and Chevalier (2003) argued that the optical afterglow evolution (the initial slow decay followed by a steepening at $\sim 0.1$ days) is consistent with the source interacting with a wind. As Wolf–Rayet winds are thought to be clumpy, the clumpiness may provide an explanation for the prominent bump in the optical light-curve (Nakar, Piran, & Granot 2003, Heyl & Perna 2003). In fact this question of wind, clumpy or homogeneous medium is difficult to fix. For instance Nakar, Piran, and Granot (2003) showed that different models, with variable-density circum-burst medium or variable energy, may reasonably fit the observed light-curve of the optical afterglow of GRB 021004. Moreover, for different values of $p$, different density profiles or energy patterns are possible. For instance, a variable external density could arise from a clumpy interstellar medium or from a variable stellar wind, while a variable energy could arise from refreshed shocks or from an angular dependent jet structure (patchy shell). The authors feel that the patchy shell model with $p = 2.2$ is nevertheless the most likely scenario (see also Schaefer et al. 2003). For GRB 030329, Gorosabel et al. (2006) reversed the argument. They observed a lack of short timescale fluctuations in the optical afterglow of this burst, which allowed them to put strong constraints on possible small-scale angular inhomogeneities within the jet and on possible fluctuations of the external density.

Wang and Loeb (2000) considered the interstellar turbulence and its influence on the light-curves of afterglow. This turbulence creates inhomogeneities in the circum-burst medium. The authors found that linear density fluctuations with $\delta n/n \lesssim 1$ on a length scale of $1-10^3$ AU could induce afterglow flux fluctuations with a fractional amplitude of up to $\sim 40\%$ over timescales of tens of seconds in the X-rays, or up to $30\%$ over tens of minutes in the optical, and of up to $9\%$ over days in the radio. Nevertheless these fluctuations average out over the range of density inhomogeneities within the emitting region and are attenuated. In addition, this variability can be masked by the variability associated with internal shells within the fireball that are catching up with the decelerating blastwave. This contamination is expected at high (X-rays) and low (radio) frequencies. Finally, the flux might scintillate at low radio frequencies due to inhomogeneities in the Galactic interstellar medium. Wang and Loeb (2000) thought that the optical–infrared domain is the best adapted to observe the density fluctuations they have considered. For a particularly well-studied event, GRB990510, their model provides interesting upper limits on the density fluctuations on a scale of 20–200 AU around the source. They are lower than the fluctuations observed on similar scales in the local interstellar medium.

After these examples illustrating how density inhomogeneities can produce temporal fluctuations in GRB afterglows, we will now turn our attention to another way of producing fluctuations: variable energy injection. Much work has been done on this question to explain the observations of the early X-ray afterglow with Swift. The most recent results, based on Swift observations, will be discussed in Chapter 7.
Section 6.9 concentrates on the early work based on BeppoSAX and HETE-2 GRBs.

6.9 VARIABLE ENERGY INJECTION

In the standard model the ejecta was treated as a uniform sphere or shell where most of the energy is carried by material with a well defined and generally high Lorentz factor. This is the simple uniform jet model with a single value of $\Gamma$ and $E$.

More physically realistic is a situation where the ejecta has a range of bulk Lorentz factors, with larger amounts of mass and energy having lower bulk Lorentz factors (Rees & Mészáros 1998). The outer shock and contact discontinuity decelerate as the fireball sweeps up external matter. This deceleration allows slower ejecta to catch up with the forward shock, replenishing and re-energizing it, and it leads to additional dissipation in the reverse shock. The energy available to power the afterglow may substantially exceed that of the burst itself. Such models can produce very slow decay rates in the afterglow, they can accommodate very different ratios of afterglow to burst fluences and they can explain different afterglow light-curves (Rees & Mészáros 1998). Rees and Mészáros (1998) and others (Panaitescu, Mészáros & Rees 1998, Sari & Mészáros 2000) have considered a fireball in which the mass fraction ejected above a given initial bulk Lorentz factor decreases according to $M(> \Gamma_f) \propto \Gamma_f^{-s}$. For $s > 0$ the mass is dominated by shells with low $\Gamma_f$, while the energy is dominated by shells with low $\Gamma_f$ for $s > 1$. Of course, other distributions of $\Gamma$ and energy can be considered. For instance, if the fireball is modeled by two shells with very different $\Gamma_f$ with the slower shell carrying more energy, the afterglow luminosity would initially decline as in the standard monoenergetic case but when the blastwave has slowed down enough for the slower shell to catch up, there would be a boost of the luminosity before resuming the power-law decline. Changes in the slope of the flux decay or late bumps can be explained in this way (Rees & Mészáros 1998). More generally a distribution of Lorentz factors that is discontinuous can explain sudden increases in the flux of the afterglow, leading to bumps in the light-curve. This possibility was used by Granot, Nakar, and Piran (2003) to explain two properties of GRB 030329: a low-energy output in gamma rays, combined with a large bump in its optical afterglow after 1–2 days.

This ‘refreshed shock scenario’ can be based on slow, massive shells and fast shells emitted simultaneously or on shells with similar Lorentz factors ejected at different times. The energization of the shock can be carried out in discrete episodes or continuously. A continuous injection of energy with a lower Lorentz factor tends to flatten the decay slope of the afterglow after it has gone through the maximum. On the other hand, an external density gradient (as in a wind) has the property of steepening the decay. A test of this continuously refreshed scenario is that it predicts a spectral maximum in the far infrared or millimeter range after a few days (Sari & Mészáros 2000). As an example Björnsson et al. (2002) have interpreted the optical afterglow of GRB 010222, one with the slowest decay, as being due to a jet with a continuous energy injection. This avoids having either an electron energy distribution
with a power-law index \( p < 2 \) or a quasi-spherical outflow with the usual energy problem. This continuous energy injection allows a value of \( p \) around 2.5, fully consistent with what is expected from modeling the acceleration in relativistic shocks (\( p \approx 2.2–2.3 \); Achterberg et al. 2001). Zhang and Mészáros (2002b) have also considered continuous energy injection, either in the form of a Poynting-flux-dominated outflow, or as a kinetic-energy-dominated matter shell injected after the burst. They calculated the corresponding afterglow light-curves in the X-ray, optical and radio bands (Figure 6.20).

Continuous energy injection can explain the optical afterglows of several events which present bumps in their light-curves, such as GRB 970228, GRB 970508, GRB990326. These afterglows can also be explained with other interpretations: refreshed shock (Panaitescu, Mészáros, & Rees 1998), a supernova component (Bloom et al. 1999, Galama et al. 2000), dust echoes (Esin & Blandford 2000), gravitational microlensing events (Garnavich, Loeb, & Stanek 2000). Zhang & Mészáros (2002b) discussed how to discriminate among these different interpretations.

In fact, the fluctuations observed in the light-curves of GRB afterglows can either be due to irregular density profiles or to variable energy injection. These two cases can be treated by the general formalism proposed by Nakar and Piran (2003), which gives a good approximation of the observed afterglow when the external density and/or blast wave energy varies, as long as the variations are spherically symmetric. These authors applied their calculations successfully to the peculiar afterglow of GRB 021004, which showed clear deviations from a smooth power-law decay. They found that the steep decay after the first bump could not result from spherically symmetric density variations nor from the passage of \( \nu_{\text{BL}} \) through the optical band. This led them to suggest that an angular structure within the ejecta or within the external medium could be present in the early stages of the afterglow of GRB 021004 (Nakar & Piran 2003, Nakar, Piran, & Granot 2003).

To summarize, the variability observed in some afterglow light-curves may be due to different scenarios:

- The ambient density into which a blast wave expands may have fluctuations.
- The emitting surface may have an intrinsic angular structure. This is the patchy shell model, which has been discussed by Mészáros, Rees, and Wijers (1998) and Kumar and Piran (2000a).
- The shock can be refreshed by slow shells with lower \( \Gamma \) catching up with the decelerated leading shells, which were initially faster.

The central engine may still be active during the afterglow. This has been proposed by Ioka, Kobayashi, and Zhang (2005) to explain the afterglows of GRB 011211 and GRB 021004. The compact object powering the jet may continue to eject an intermittent outflow on a very long timescale (>1 day). This is the opportunity to report a proposal by Rees and Mészáros (2000): the possibility of explaining all afterglows by the activity of the central engine itself (in this case a magnetar). For
Figure 6.20.
Broadband signature (X-ray, optical, radio) of the injection of a kinetic-energy-dominated shell. Solid lines are light-curves with energy injection compared to dashed lines without injection. Part (a) shows the light-curves expected after a violent injection involving formation of new shocks. Part (b) shows the light-curves expected after a mild injection (Zhang & Mészáros 2002b).
a more complete discussion of variability in GRB afterglows see Ioka, Kobayashi, and Zhang (2005).

6.10 THE FIREBALL MODEL CONFRONTS OBSERVATIONS

The basic standard model that we have described was applied with remarkable success to the first GRB afterglows discovered with BeppoSAX. Afterglow spectra and their light-curves at various wavelengths are correctly described by the emission of an adiabatically expanding blastwave radiating its energy when it decelerates during its progression into the surrounding medium. The ambient density and the energy of the blastwave can be calculated. The energy densities in electrons and in the magnetic field could also be computed, for instance for the famous GRB 970508, with respective values of 12% and 9% of the nucleon energy density, not very far from the equipartition as assumed in the simplified model.

Already by the time of the first BeppoSAX bursts Panaitescu, Mészáros and Rees (1998) had considered modifications of the simple model, like a delayed energy injection within an axially symmetric jet. This concept of a jet was also put forward, for instance, by Sari, Piran, and Halpern (1999), in the analysis of GRB 980519 which displayed very rapid fading at optical and X-ray wavelengths ($t^{-\alpha}$ with $\alpha = 2.05 \pm 0.04$). Beaming was also seen as a good way to relax the energy requirements on some of the most energetic GRBs. Kulkarni et al. (1999a) estimated the isotropic energy released by GRB 990123 at redshift $z = 1.6$ to be $3.4 \times 10^{54}$ erg $\sim 1.9 M_\odot c^2$, more than the rest mass energy of a neutron star. They argued that such energies require a beaming of the GRB emission. The presence of a jet was also reported for instance by Harrison et al. (1999) and Stanek et al. (1999) for GRB 990510. The simple jet model introduced by these authors led to a jet opening angle $\theta_0 \sim 0.08 (n/1 \text{ cm}^{-3})^{1/8}$, about $5^\circ$, reducing the isotropic gamma-ray energy release by a factor of $\sim 300$, from $2.9 \times 10^{51}$ erg to $10^{51}$.

The effect of the density of the circum-burst medium was also considered very early. Kumar (2000), for instance, compared on a statistical basis the distribution of GRB afterglow fluxes with afterglow models invoking a uniform interstellar medium or a stratified medium (a stellar wind). Chevalier and Li (1999, 2000) made this comparison for individual GRBs and concluded that GRB 990510 and GRB 990123 could be better explained by a constant density medium than by a wind-density profile. On the other hand, Li and Chevalier (2003) argued that the early afterglow of GRB 021004 was consistent with the fireball interacting with a Wolf–Rayet wind. Scalo and Wheeler (2001) emphasized that the massive star winds and supernovae expected in a cluster of massive stars might create large superbubbles with low densities. The variation in the cluster ages and in the density of giant molecular clouds in the superbubble might explain the quite large dispersion in the reported densities of the circum-burst medium. Chevalier, Li, and Fransson (2004) investigated the possibility that the fireball expands in the shocked wind bubble in a starburst region and suggested that the surroundings of GRBs could have been modified by stellar winds or ionizing radiations since high densities
 (>10^3 cm\(^{-3}\)) are not observed. These comments demonstrate why the problem of the nature and the density of the circum-burst medium and of the jet structure are still quite open issues.

The modeling of the multi-wavelength emission of the afterglows of a growing population of GRBs allowed Panaitescu and Kumar (2001a, 2002) to determine the physical parameters of GRBs: the kinetic energy of the ejecta after the prompt GRB phase, the initial jet aperture, the density of the ambient medium, and the slope of the power-law electron distribution (which does not appear as universal as assumed in the simple version of the fireball model). They found the following values: a burst kinetic energy \(\sim 5 \times 10^{50} \text{ erg}\), an initial jet aperture between 2° and 20°, half of the jets being narrower than 3°, and a density of the circum-burst medium varying between 0.1 and 50 cm\(^{-3}\), with a preference for a constant density interstellar medium. At the deceleration radius \(\Gamma\) has reached values between 70 and 300, and the mass of the ejecta is evaluated around \(10^{-6} \, M_\odot\). These analyses demonstrated the importance of multi-wavelength afterglow data to characterize the ejecta and its behavior into the surrounding medium. Concerning the microphysics of the shocks the authors found that:

(a) The total electron energy is close to equipartition with protons.
(b) The strength of the magnetic field, \(B \propto \varepsilon_B^{1/2}\), is within two orders of magnitude of the equipartition value.
(c) The slope, \(p\), of the power-law distribution of the shock-accelerated electrons is not universal, and \(p\) is found to range between 1.4 and 2.8.

But as the authors noticed, several assumptions have been made: a uniform jet, the energy parameters \(\varepsilon_e, \varepsilon_B\) are assumed to be constant, and the observer is on the jet axis. Yost et al. (2003) have undertaken the same kind of study using the fireball model for four GRBs. They have considered the basic afterglow model as well as deviations from it. First, they proposed to see if a variable \(\varepsilon_B\) was allowed by the data. As any change in the magnetic energy fraction is expected to come from the evolution of the shocks they consider \(\varepsilon_B\) being tied to the shock strength expressed by its Lorentz factor \(\Gamma\), so they considered \(\varepsilon_B \propto \Gamma^\alpha\), the basic model having \(\alpha = 0\). The conclusions obtained by Yost et al. (2003) are quite interesting: the data do not constrain the fireball model to have a constant magnetic energy fraction \(\varepsilon_B\). Good fits are possible with \(\varepsilon_B\) both increasing and decreasing with \(\Gamma\), variations like \(\varepsilon_B \propto \Gamma\) through \(\varepsilon_B \propto \Gamma^{-2}\) are permitted. With \(\Gamma\) changing by a factor of \(\sim 10\) over the data range this allows an extra change in magnetic field strength by a factor of 3–10. The diversity in the fitted microphysical parameter values should be noted. The magnetic energy fraction varies from 0.2% to 25% (two orders of magnitude!). A constant density gives the better fit with \(n \sim 0.2 \, \text{cm}^{-3}\) for one burst, and \(n \sim 20 \, \text{cm}^{-3}\) for the three other bursts. The collimation varies from near-isotropy to a half-angle of 0.04 radians. The kinetic energy is comparable to the total GRB \(\gamma\)-ray energy. The main conclusion is that the model assumptions are not strongly constrained by the data even when these data are good-quality multi-wavelength data. Another important question which is still pending concerns the power-law index of the
electron distribution, \( p \), which can vary from \( p = 2.1 \) to 2.9. On-going work strongly suggests that there is not a universal value of \( p \), but rather a distribution of spectral indices.

Another aspect, which has been considered extensively by many authors, concerns the effects inverse Compton (IC) radiation losses. Wu et al. (2005b) noted that IC scattering plays an important role when the energy equipartition factor \((\varepsilon_E)\) of the electrons is much larger than that of the magnetic field \((\varepsilon_B)\). IC radiation prolongs the fast-cooling phase during which the relativistic shock is semi-radiative and stays semi-radiative after the transition into the slow-cooling phase. The authors computed the temporal decay index of the X-ray afterglow in this semi-radiative phase. Considering \( p = 2.2–2.3 \) and \( \varepsilon_E \sim 1/3 \), they found a temporal index of the X-ray light-curve, \( \alpha_x \), that is more consistent with the observed value \( \alpha_x \sim 1.3 \) than for a purely adiabatic model (Section 6.4). They also compared the multi-wavelength afterglows expected from semi-radiative blastwaves with the adiabatic afterglow light-curves obtained in the synchrotron-dominated cooling regime.

In conclusion, the pre-Swift era has provided strong support for the fireball model, but various issues remained:

- The problem of the circum-burst medium. Why, with a collapsar as progenitor, does a homogeneous circum-burst medium provide better fits to the afterglows than a stratified wind?
- Are the breaks due to a jet or to another process (for instance, cooling breaks or the cessation of energy injection in the forward shock)?
- Are the reported breaks really achromatic?
- What is the beam configuration, homogeneous or structured jet?
- How can the range of \( p \) values inferred from the observations be explained?
- Are microphysical parameters constant during the development of the afterglow?
- Can \( \varepsilon_B \) be a variable?

As noted by Mészáros (2006), the microphysics of particle acceleration, magnetic field amplification in shocks and/or reconnection or dissipation are far from being well understood. For the prompt GRB phase the questions of efficiency and of the radiation mechanisms are not clarified. Another question is the role of the reverse shock and the small number of bright optical flashes. The answers to all these questions might lead to complications of the simple external–internal shock model. This is also the opportunity to say that the nature of the progenitor and of the underlying central engine are still poorly known.

### 6.11 Conclusion

In this chapter, we have analyzed the production of the afterglow with the fast expansion of the fireball, the shock on the surrounding medium, and the slowing down of the shells. The afterglow emission starts when most of the energy of the ejecta is transferred to the shocked external medium. In this phase the shocked region
at a distance $R$ progresses in the external medium and its Lorentz factor decreases with $R$. Knowing $R(t), \Gamma(t)$ and the shock conditions, the energy spectrum and the light-curves of the afterglow have been calculated. The observed emission is described globally by $F_\nu \sim t^{-\alpha} \nu^{-\beta}$. $F_\nu \sim t^{-\alpha}$ allows analysis of the decaying part of the light-curve at a given frequency $\nu$, while $F_\nu \sim \nu^{-\beta}$ allows characterization of the afterglow spectrum at a given time. The synchrotron spectrum has been represented by segments of power laws between characteristic frequencies $\nu_{\text{sa}}, \nu_c, \nu_{\text{mb}}$; on these segments the indices of the power law are constant or related to the power-law index of the electron spectrum. Different relations between $\alpha$ and $\beta$, called closure relations, have been obtained for various conditions such as adiabatic or radiative situations, fast or slow cooling or different environments. The comparison of the theoretical values and those deduced from the observations, light-curves and spectra at different wavelengths, have been used to confirm the shock model.

In 1997 after the first afterglow observations of several GRBs by BeppoSAX the comparisons with the standard model were very encouraging in spite of the simplifications of the model. However, with the accumulation of afterglow observations, it became evident that the model was too simple. More complex conditions were introduced very soon after the discoveries of BeppoSAX. They have been analyzed in the second part of this Chapter 6 and they include:

- A reverse shock as an explanation of the strong optical flashes and radio flares; some examples have been discussed indicating its possible role, but there are other explanations such as internal shocks or a forward shock at its early stage. Moreover as few optical flashes have been detected we have tried to answer the following question: Why are there few bright early optical flashes?
- Non-spherical relativistic ejecta. GRB outflows have been considered as possibly collimated into narrow jets. In this case signs of the presence of jets have been looked for. In the BeppoSAX era the observation of achromatic breaks in the afterglow light-curves of several GRBs was considered to be an indirect evidence of jets. But we have seen that the presence of jet breaks in the afterglows can be difficult to find because the break might be absent, or very weak, or at a very early or very late time. Moreover hydrodynamic simulations have indicated that the lateral expansion of the jet is modest. Hence there are several reasons which can explain the absence of jet breaks in many bursts, even if the ejecta is collimated as is demanded for some GRBs whose isotropic energy $E_{\text{iso}}$ is dramatically large. We will see in Chapter 7 how the absence of achromaticity of the temporal breaks is a major concern.
- In spite of these recent doubts cast on the breaks and their meaning, they have been used to determine the beaming of the jet. The measurement of the time of the break makes it possible to obtain $\theta_j$, the beaming angle of the GRB, and the true energy output when the redshift is determined. We have discussed the determination of the $\gamma$-ray energy release and its comparison with the kinetic energy remaining in the ejecta. We have also analysed their connection with the total energy of the relativistic ejecta emitted by the central engine. At the time of BeppoSAX these analyses indicated that the energy output of the central engine
seemed to take on a universal value for long-duration GRBs and that the typical opening angles for the collimated bursts lie between 1° and 20°.

- The presence of jets associated with GRBs has led to the determination of the beaming factor and to the conclusion that a vast majority of GRB/jets are not directed towards the Earth and therefore that many GRBs must be seen off-axis. This means that there exist afterglows without GRBs. They have been called orphan afterglows and we have reported on the searches for orphan afterglows at X-ray, visible, and radio wavelengths. Such studies are useful to establish the presence of jets and to get an idea of the beaming factor independently of jet breaks. Today these searches have revealed no orphan afterglow candidate.

- If the relativistic ejecta seem to be collimated into narrow jets a complementary question is to know how the energy is distributed inside the jet. We have discussed different types of jets: uniform, and universal structured jets, which are among the most popular models, and evoked other models such as jets with a Gaussian angular profile or two-component jets. We have seen that it is extremely difficult to infer the structure of the jet from the light-curves of the afterglows.

- The polarization of the GRB and its afterglow has been considered. These measurements have the potential to constrain jet models and to estimate the configuration of the magnetic field in GRB ejecta. Up to now these studies have not made it possible to answer these fundamental questions.

- In the simplest model the external medium surrounding the GRB is homogeneous. But very early the possibility of an inhomogeneous external medium was introduced. The case of a wind was particularly attractive because the progenitors of long GRBs have been associated with the deaths of massive stars of Wolf–Rayet type. For these stars the presence of strong winds has been well established and therefore winds are expected in the vicinity of the progenitors of long GRBs (see Chapter 8). But even if such winds are expected it has been reported that for a majority of GRBs a homogeneous medium provides the best fits to the observations.

- The case of inhomogeneities in the circum-burst medium has also been considered: irregular density profiles, clumps of matter, and inhomogeneities due to interstellar turbulence. Introducing such density inhomogeneities has one major objective: the explanation of various temporal fluctuations in the light-curves of GRB afterglows.

- Another way to explain these temporal fluctuations is to complete the standard model by introducing variable energy injection. The expanding ejecta is no longer characterized by a single value for $\Gamma$ and $E$, but includes a range of bulk Lorentz factors allowing the duration of energy injection to be extended. Another way to have this refreshed shock scenario is to assume that the ejecta are constituted of shells with similar Lorentz factors ejected at different times. Like irregular density profiles, variable energy injection can be the source of the variability observed in afterglow light-curves. We will see in Chapter 7 that the early X-ray afterglow and the presence of X-ray flares provide strong arguments in favor of delayed energy injection, which can take place until many hours after the GRB.
But such delayed energy injection implies a long-lasting activity of the progenitors, which may be a concern.

- We have finally addressed the question of the variability of the microphysical parameters $\varepsilon_e$ and $\varepsilon_B$, which are assumed to be constant in the standard model. Their possible variability has been considered with the result that great flexibility is allowed, still producing acceptable fits.

- Finally among other critical issues we have evoked the importance of inverse Compton depending on the values of the energy equipartition parameters $\varepsilon_e$ and $\varepsilon_B$.

Clearly in the pre-Swift era, in spite of the development of more complex fireball models, many questions were not definitively answered: the nature of the circumburst medium, the meaning of the observed breaks in the afterglow light-curves, the jet structure, the universality of $p$, and the possible variability of $\varepsilon_e$ and $\varepsilon_B$ and their consequences. In Chapter 7 we will see that the Swift era has led to discoveries of the highest importance, which instead of solving these issues made most of them even more crucial.

6.12 REFERENCES


Early afterglows: the Swift revolution

In Chapter 4 we presented the discovery and characterization of GRB early afterglows by Swift. As we have seen, this was a major breakthrough, which contributed to a better knowledge of GRBs. In particular, these observations contributed to a better appreciation of the connection between the GRB prompt emission and the afterglow. Swift has brought various surprises: the complex behavior of the early X-ray afterglow (Nousek et al. 2006, O’Brien et al. 2006), the frequent faintness of the early optical afterglow (Roming et al. 2006), and a general lack of synchronism of X-ray breaks and optical breaks challenging the jet break interpretation (e.g. Panaitescu et al. 2006a, Panaitescu 2007a,b).

The analysis of the early X-ray afterglow data, in connection with those of the prompt gamma-ray emission and of the early optical/IR afterglow, is crucial for the study of many questions, among them:

- Does the central engine activity stop after the burst or does it last longer with late injection of matter and/or magnetic fields?
- What is the immediate environment of the burst: the interstellar medium or a wind?
- What is the role of inhomogeneities in the circum-burst medium?
- What is the role of the reverse shock?
- Can we get a better approximation of the initial Lorentz factor of the fireball?
- What is the impact of a possible temporal/spatial dependence of the microphysical parameters, $\varepsilon_e$ and $\varepsilon_B$?
- What is the slope $p$ of the spectrum of the electrons accelerated in the shocks?
  Does it vary? Is there a break in the power-law electron spectrum?

Swift, with its capacity to detect an average of two GRBs per week and to quickly point its narrow-field instruments towards them, is a formidable tool to respond to many of these open issues. In particular, the XRT, a sensitive imager in the
0.2–10 keV energy range, is ideal for studying the transition between the GRB prompt emission and the very early X-ray afterglow. Fast, multi-wavelength follow-up from the ground is the other necessary ingredient for understanding the complex physics at work in early GRB afterglows.

The contribution of Swift in this field has been decisive. We have presented in Chapter 4 the main characteristics of the early X-ray light-curves. Figure 4.13 shows a cartoon of the X-ray light-curve. It allows the identification of the main phases following the prompt emission: a steep decay which appears between 10 s and a few hundred seconds, a shallow or slow decay phase which can last much longer, typically between $10^3$ s and few $\times 10^4$ s, and the other two phases already reported in the pre-Swift period: the pre-jet-break and the post-jet-break phases. In most cases the pre-jet-break phase ends during the first day but it can last for 10 days in some GRBs. The flux in this phase has a time-dependence $t^{-\alpha}$ with $\alpha \sim 1.2$, consistent with the prediction of the standard afterglow model. The post-jet-break phase appears after what has been called the jet break. It is difficult to cover this phase with the XRT because it usually appears at late times, sometimes several days after trigger, when the X-ray afterglow is already very faint or undetectable. A typical X-ray light-curve is well illustrated by GRB 050315. Figure 7.1 describes the different phases of its afterglow with the characteristic times and the values $\alpha$ and $\beta$ for each phase (from Panaitescu 2006).

In a sample of 40 GRBs O’Brien et al. (2006) found that the initial X-ray emission seen in the XRT is usually consistent with being the extrapolation at low energies of

![Figure 7.1](image)

**Figure 7.1.** The phases of the afterglow of GRB 050315. This GRB provides a good illustration of the typical times and properties of each phase. The figure indicates the characteristic values of $\alpha$ (the temporal decay index) and $\beta$ (the spectral index) of each phase (from Panaitescu 2006).
the emission seen by BAT. Among them, two-thirds of the GRBs display the typical light-curve discussed above, while the remaining bursts decline fairly continuously.

At least half of the GRBs display X-ray flares (see Chapter 4), but only for a few of them is the fluence in the flares comparable with that of the burst. Flares typically happen hundreds of seconds after the trigger or earlier, but in some cases they occur much later, around 1 day. For a sample of 37 afterglows observed mostly with Swift, Panaitescu (2006) reported 19 flaring X-ray afterglows with two-thirds exhibiting flares during the fast decay phase. The steep decay and the flares are considered by Panaitescu (2006) as originating from internal shocks (see Sections 7.1.1 and 7.4.2).

To summarize, for this early afterglow phase the major discoveries of Swift are: the steep decay phase at the end of the prompt emission, the existence of a plateau (the shallow phase) preceding the standard afterglow decay, and the presence of important X-ray flaring activity. Zhang et al. (2006), taking into account these new results, analyzed the physical processes shaping GRB X-ray afterglow light-curves and their theoretical implications. Their very complete review is the baseline for our analysis. We will now consider successively the steep decay, the shallow phase, and the X-ray flares. We will not return to the last two phases, pre- and post-jet-break, which have been covered already in the pre-Swift period. Finally we will see how the chromatic breaks in the X-ray light-curve can be explained, and this will be the opportunity to return to the crucial issue of the jet breaks, already discussed in the pre-Swift era.

7.1 THE STEEP DECAY AND ITS ORIGIN

The steep decay phase was identified as a common feature of the early X-ray afterglows very soon after the launch of Swift (Tagliaferri et al. 2005). It is commonly believed that the prompt GRB emission originates from some processes taking place within the ejecta, the $\gamma$-rays being emitted in internal shocks before the fireball is decelerated by the circum-burst medium. The afterglow occurs when the fireball is decelerated, and its emission appears at a much greater distance from the central engine. Therefore, the prompt emission and the afterglow seem to arise from different emission sites. If the flux level of the prompt emission is much higher than the afterglow emission flux level we expect to see a steeply decaying light-curve during the transition from the prompt emission phase to the afterglow phase. In Section 7.1.1 we will see that such a steep decay is explained by the so-called curvature effect (Kumar & Panaitescu 2000, Dermer 2004, Panaitescu et al. 2006b). So this steep emission component can be regarded as the tail of the prompt GRB emission.

7.1.1 The curvature effect

For a conical jet with an opening angle $\theta_j$, the emission from the same radius $R$ but from different viewing latitudes $\theta$ would reach the observer with some delay. Therefore, even if the emission stops instantaneously, due to the propagation effect
the observer would receive the emitted photons at an angle $\theta$ with a delay $t = (1 + z)(R/c)(\theta^2/2)$, where $z$ is the redshift of the GRB. As the emission is limited by $\theta_j$, the tail emission corresponding to the time delay of the emission from $\theta = 0$ to $\theta = \theta_j$ will have a duration of the order of $t_{\text{tail}} = (1 + z)(R/c)(\theta_j^2/2)$. This assumes that the line of sight is not too close to the jet edge. The electrons are assumed to be shock-heated up to the radius $R$ with no shocked electrons injected beyond radius $R$. The comoving emission frequency $\nu'$ of the shock-heated electrons is transformed to $\nu = D\nu'$ in the observer frame (where $D = [\Gamma(1 - v \cos \theta/c)]^{-1}$ is the Doppler factor).

$D \sim 2/\Gamma \theta^2$ for $\theta \gg 1/\Gamma$, and since we have seen that $t \propto \theta^2$ one obtains $D \propto t^{-1}$. The observed flux $F_\nu$ is related to the comoving surface brightness $L'_\nu$, (see, for instance, Zhang et al. 2006)

$$F_\nu \propto L'_\nu D^2 \propto \nu^{-\beta} D^2 \propto \nu^{-\beta} t^{-2-\beta}$$

$\beta$ is the observed spectral index of the photon spectrum around the observed frequency $\nu$. This proportionality is valid for $1/\Gamma \ll \theta < \theta_j$.

The curvature effect introduces a relation between the temporal index, $\alpha$, and the spectral index, $\beta$: $\alpha = 2 + \beta$ (Zhang et al. 2006, Kumar & Panaitescu 2000, Dermer 2004, and Panaitescu et al. 2006b). This prediction of a relation between the temporal index $\alpha$ and the spectral index $\beta$ can be used to verify that the steep decay phase following the prompt emission is due to the curvature effect.

A sample of 28 GRBs was used by Panaitescu (2006) to check if the relation $\alpha = 2 + \beta$ holds. He finds that for more than half of the bursts, the fast decay satisfies the relations whereas a faster fall-off is found for a quarter of the sample and few afterglows exhibit a slower decline. Similar conclusions have been reached by O’Brien et al. (2006) for 40 GRBs:

‘Comparing the temporal and spectral indices of the power law decline, the distribution is consistent with a simple model in which the early emission is a combination of emission from the central engine (parameterised by high latitude emission) and afterglow’.

and by Nousek et al. (2006) for 27 GRBs:

‘We find that the most promising explanation for the initial fast flux decay is that it is the tail of the prompt gamma-ray emission which is emitted from large angles ($\theta > \Gamma^{-1}$) relative to our line of sight (Kumar & Panaitescu 2000). This model produces a sharp flux decay with $\alpha = 2 + \beta$, in rough agreement with observations’.

It should be noted that this interpretation of the fast decay phase strongly supports the internal origin of the prompt emission and X-ray flares. This may be taken as an observational proof that the GRB prompt emission is coming from a different site than the afterglow emission. The energy dissipation region for the prompt emission would be well inside the region where the deceleration of the whole fireball occurs.
Despite the consensus on the interpretation of the steep decay, a significant fraction of the bursts (about 40%) exhibit a temporal decay which is flatter or steeper than predicted by the curvature effect. This led Zhang et al. (2006) to examine various ways of modifying the slope of the tail of the prompt emission.

To produce a flatter decay, the following effects have been invoked. (1) An early contribution of the forward shock emission, appearing near the end of the tail emission, can change the temporal slope of the steep decay phase. This happens if the fireball is already decelerated at the end of the tail emission and if the forward shock emission contributes to the emission in the X-ray band. (2) The central engine activity may not die abruptly, or the shocked region may not cool abruptly but decay with time gradually (Fan & Wei 2005, Zhang et al. 2006).

To produce slopes steeper than 2 + \( \beta \), Zhang et al. (2006) showed that the solid angle of the emitting region must be comparable to or smaller than 1/\( \Gamma \). This would correspond to a patchy shell (Kumar & Piran 2000) or a mini-jet (Yamazaki et al. 2004). But this assumes a very particular choice for the line of sight, which seems to have a low probability.

### 7.1.2 The time zero point \( t_0 \) effect

In general, \( t_0 = 0 \) is defined by the trigger time of the burst. There is no problem with this definition for the study of late afterglows. When discussing the very early afterglow and its connection to the prompt emission, the decay power-law index \( \alpha \) is very sensitive to the value of \( t_0 \). So the choice of \( t_0 \) is crucial (see Kobayashi & Zhang (2007) for a detailed analysis of this issue). Figure 7.2 (from Zhang et al. 2006) illustrates the effect of \( t_0 \) on the light-curves. Keeping \( t_0 \) as the beginning of the first pulse (i.e. the GRB trigger) leads to a very steep power-law decay for later pulses. But it is clear that in a GRB with several pulses or peaks, each pulse corresponding to an ejected shell is followed by its tail emission due to the curvature effect but this emission is buried under the main emission of the following pulse. Therefore, one has to consider that the last pulse in the GRB largely dominates the tail emission. In this situation \( t_0 \) has to be associated to the time of this last pulse and not to the trigger time of the burst. Liang et al. (2006), with a sample of 22 bursts, made another analysis. They considered that the decay slope should be 2 + \( \beta \) and they searched for the appropriate \( t_0 \) which allows this constraint to be respected. They found that under these conditions \( t_0 \) is usually at the beginning of the last pulse, or at the beginning of the X-ray flare for steep decay phases following flares.

Various other possibilities to modify the ‘curvature relation’ have been discussed. Zhang et al. (2006) looked at the effect of a decelerating outflow, since the Lorentz factor of the shell might be decreasing before the sudden cessation of the emission. They calculated the curvature effect for a sudden switch-off of the radiation in a decelerating outflow. They found that the curvature relation is essentially unchanged due to this effect. A second possibility which might modify this relation is the presence of a jet structure. Again Zhang et al. (2006) concluded that the curvature effect is largely insensitive to the jet structure, as long as the viewing angle is not too far off the bright beam. This is also the conclusion of Dyks, Zhang, and Fan (2005) who...
indicated that light-curves with a slope shallower than predicted could nevertheless be produced for viewing angles larger than the core of the jet.

### 7.1.3 Other effects

Another explanation of the steep decay phase has been proposed by Pe’er, Mészáros, and Rees (2006). It is based on a hot plasma ‘cocoon’ associated with the jet, which expands relativistically after the jet has broken through the stellar envelope. Multiple Compton scattering inside the cocoon causes a spread in the arrival time of the X-ray photons. Calculations gave steep decays, which resemble those observed in many GRBs.

Zhang, Liang, and Zhang (2007) performed a systematic, time-dependent, spectral analysis of 44 steep decay tails. For 11 GRBs no spectral evolution is observed, but 33 of them display spectral evolution. The curvature effect alone cannot explain these results. Zhang, Liang, and Zhang (2007) suggested that spectral evolution in the
steep decay tails might be due to an evolving exponential spectrum, while a simple superposition of the curvature effect and an underlying soft decaying afterglow cannot explain the observations. Yonetoku et al. (2008) observed a strong spectral evolution during the fast decay tail of GRB 060904A. They suggested that it was due to the detection of the high-energy cut-off of the synchrotron radiation (the synchrotron frequency of the maximum energy electrons). They noted that inefficient acceleration of the electrons is needed during this phase for the high-energy cut-off to appear at X-ray energies.

After the steep phase the emission continues to fade, indicating that the fireball has already been decelerated at the time of the end of the prompt emission. Otherwise one would see an initially rising light-curve peaking at the fireball deceleration time. It can be concluded that \( t_{\text{dec}} \) arrives before the end of the steep tail, providing a lower limit on \( \Gamma_0 \), the initial Lorentz factor of the fireball.

From the relations given in Chapter 6, we have:

\[
R_{\text{dec}} = \left( 3E_0 / 4\pi n_{\text{ext}} m_p c^2 \Gamma_0^2 \right)^{1/3} \quad \text{and} \quad t_{\text{dec}} = R_{\text{dec}} / 2c \Gamma_0^2
\]

where \( E_0 \) is the outflow energy (\( E_0 = E_{\text{iso}} / \varepsilon \), where \( \varepsilon \) is the radiative efficiency) and \( \Gamma_0 = \eta = M_0 / M_{\text{ext}} \) is the terminal bulk Lorentz factor.

\[
\Gamma_0(t_{\text{dec}}) = \left( 3E_{\text{iso}}(1 + z)^3 / 32\pi n_{\text{ext}} m_p c^5 \varepsilon t_{\text{dec}}^3 \right)^{1/8}
\]

Zhang et al. (2006) used the end of the steep tail as an upper limit on \( t_{\text{dec}} \), allowing them to determine a lower limit on \( \Gamma_0 \). Knowing the redshift and the isotropic \( \gamma \)-ray energy they found values of \( \Gamma_0 \) ranging from 100 to 200(\( \varepsilon n_{\text{ext}} \))\(^{-1/8} \), with a weak dependence on \( \varepsilon \) and \( n_{\text{ext}} \). Their conclusion is that these data suggest that GRBs are highly relativistic events with Lorentz factors higher than 100. Molinari et al. (2007) using the peak of the afterglows of GRB 060418 and GRB 060607A observed in the NIR, determined the onset of the afterglows (see their Figures 1 and 2) and were able to establish the deceleration time of the fireball. They found a Lorentz factor of the fireball \( \Gamma_0 \sim 400 \) (for both GRBs, assuming an ISM environment) and a deceleration radius \( R_{\text{dec}} \sim 10^{17} \) cm, well beyond the typical value of \( 10^{15} \) cm that we have reported for the internal shocks, and in agreement with the values given for the appearance of external shock (see Chapter 6). At the same time the baryonic load can be obtained: \( M = M_0 / \Gamma_0 = E_0 / \Gamma_0 c^2 \sim 7 \times 10^{-4} \) \( M_\odot \).

### 7.2 THE SHALLOW PHASE

During this phase (number II in Figure 4.13) the X-ray afterglow light-curve is rather flat, with a temporal decay index \( \alpha \) varying between 0.2 and 0.8, and no variation of the spectral index during the transition between the shallow to normal decay phase (the pre-jet phase). On the other hand there is evidence that the X-ray continuum sometimes hardens at the transition from the fast decay to the shallow-decay phase. Analyzing 26 X-ray afterglows, Panaitescu (2006) found that all X-ray decays in this phase are too slow for the standard forward-shock model indicating that an
additional mechanism has to be introduced to slow down the decay. We explain below that it is natural to introduce a continuous energy injection that will ‘refresh’ the forward shock to explain this shallow-decay phase.

In this shallow phase, the X-ray flux $F_X \propto t^{-\alpha}$, with $0.2 < \alpha < 0.8$, and $t F_X(t)$ increases with time. If we call $\varepsilon_X(t)$ the efficiency of the X-ray afterglow emission, $E_{K,iso}(t)$, the isotropic equivalent kinetic energy in the afterglow measured at time $t$, and $L_{X,iso}$ the X-ray luminosity at time $t$ (we have seen that $t = 10\,h$ is widely used) we have (Granot, Königl, & Piran 2006):

$$\varepsilon_X(t) = t L_{X,iso}(t) / E_{K,iso}(t)$$

Using $L_{X,iso}(t) = 4\pi d_L^2 (1 + z)^{-\alpha - 1} F_X(t)$, this equation can also be written:

$$\varepsilon_X(t) E_{K,iso}(t) / t F_X(t) = 4\pi d_L^2 (1 + z)^{-\alpha - 1}$$

The right-hand side of the equation is constant and the equation can be written:

$$\varepsilon_X(t) E_{K,iso}(t) \propto t F_X(t)$$

This shows that $\varepsilon_X(t) E_{K,iso}(t)$ has to increase with time as $t F_X(t)$.

Given the expected decrease of $\varepsilon_X(t)$ with $t$ for $p > 2$ (see Granot, Königl, & Piran 2006), the increase of $\varepsilon_X(t) E_{K,iso}(t)$ has to be attributed to an increase in $E_{K,iso}(t)$ and to some sort of energy injection in the forward shock (Nousek et al. 2006, Panaitescu et al. 2006b, Zhang et al. 2006, Granot & Kumar 2006).

7.2.1 Refreshed shock: Energy injection

To produce a refreshed shock and energy injection Zhang et al. (2006, 2007), identified three possibilities:

1. a long-lived central engine with progressively reduced activity,
2. an instantaneous injection with a steep power-law distribution of Lorentz factors,
3. delayed energy injection in the forward shock.

1. In the first case, with a long-lasting activity of the central engine $L(t)$ behaves as: $L(t) = L_0 (t/t_b)^{-q}$. This has been analyzed by Zhang and Mészáros (2001). The energy injection might be due to the initial spin-down from a millisecond pulsar (Dai & Ly 1998b) or to continuing infall into a central black hole, Zhang and Mészáros (2001) have shown that for an adiabatic fireball the injection would modify the blastwave dynamics as long as $q < 1$. As inferred from the observations, the injection process has to be rather smooth with an effective $q$ value around 0.5. For this scenario of continuous injection, the luminosity of the central engine has to vary smoothly with time, in contrast to the erratic injection of energy observed during the GRB itself. For these authors, this requires two components: a hot fireball that leads to the prompt emission and a cold ‘Poynting’ flux that leads to the smooth energy injection. This second component could be due to the spin-down of a newly born millisecond pulsar (Dai & Ly 1998b), but for such a model $q = 0$. If a long-lived black hole
torus system with a slowing down accretion rate is invoked, \( q = 5/3 \) at later times (MacFadyen, Woosley, & Heger 2001, Janiuk et al. 2004), too steep to give an interesting injection signature.

In any case, independent of the energy source a long-lived activity of the central source requires the source to be active a very long time, up to several hours after the GRB, with a smooth temporal behavior, as we have already said. Most of the energy is injected around the end of the shallow phase, typically around \( 10^4 \) s, at the transition from the shallow decline to the regular decline. This late energy injection makes the problem of the efficiency of the energy conversion during the prompt phase even more critical because the \( \gamma \)-ray radiation now represents a very significant fraction of the initial kinetic energy (Nousek et al. 2006). This point is extensively discussed in Section 7.3.

(2) In the second possibility the central engine activity may be brief (as short as the prompt emission itself) but at the end of the prompt phase the ejecta has a range of Lorentz factors. The amount of ejected mass moving with Lorentz factors larger than \( \Gamma \) is given by (Rees & Mészáros 1998, Panaitescu, Mészáros, & Rees 1998, Sari & Mészáros 2000):

\[
M(> \Gamma) \propto \Gamma^{-s}
\]

In this scenario the activity of the central engine does not need to be long-lasting and the material can be ejected rapidly. The continuous energy injection is due to the different velocities of the ejecta. The slower ejecta progressively piles up onto the blastwave as the latter decelerates. A change in the fireball dynamics is expected for \( s > 1 \) when most of the energy is stored in the slow-moving material. For each value of \( s \) one can find an effective value of \( q \) that mimics the effect of \( s \) (or vice-versa; see eq. 20 and 21 in Zhang et al. 2006). For \( q = 0.5 \), for instance, one gets \( s = 2.6 \) for an ISM environment. The temporal break around \( 10^3-10^4 \) s suggests a cut-off of the Lorentz factor around several tens (Zhang et al. 2006).

Granot and Kumar (2006) also noticed that the end of the ‘shallow’ flux decay phase marks the beginning of the Blandford–McKee (BK) self-similar external shock evolution. They determine the time-dependence of the blastwave Lorentz factors in the framework of this model and they show that the Lorentz factor typically drops by a factor of 2–4 during the shallow decline phase (in the case of a uniform circum-burst medium). With their model based on a deviation at early times from the constant energy BK self-similar solution, they predict a roughly similar shallow decline in the optical band over the same time interval as in the X-ray data. The authors also suggest that most of the energy in the relativistic outflow is in material with Lorentz factors in the range 30–50 (see their Figure 1). In this scenario of energy injection with a wide distribution of Lorentz factors, the reverse shock is typically non-relativistic (Sari & Mészáros 2000) because the relative Lorentz factor between the injection shell and the blastwave is always low when the former piles up onto the latter.

(3) The third possibility considered by Zhang (2007a) is not related to the central engine. It concerns a delayed energy transfer to the forward shock. As explained in Chapter 6, the onset of the afterglow is estimated to be around \( t_{\text{dec}} = \max(t_\gamma, T) \),
where $t_\gamma$ is the timescale at which the fireball collects $\Gamma^{-1}$ of the rest mass of the initial fireball from the ISM and $T$ is the duration of the GRB. Numerical calculations have suggested that the timescale before entering in the Blandford–Mckee self-similar deceleration phase is long (several $10^3$ s, Kobayashi & Zhang 2007). So, it takes time to transfer kinetic energy of the fireball to the medium. Under these conditions the shallow decay phase may simply reflect the slow energy transfer from the ejecta to the ambient medium (Zhang et al. 2007). Another possibility of delayed energy transfer is obtained when the outflow has a significant fraction of Poynting flux (Thompson 1994, Mészáros & Rees 1997, Lyutikov & Blandford 2003, Zhang et al. 2006). The outflow is characterized by a parameter $\sigma$ which is the ratio between the Poynting flux and the baryonic kinetic energy flux. Zhang and Kobayashi (2005) have modeled the reverse-shock emission from ejecta with arbitrary values of $\sigma$. They found that the Poynting energy is not transferred to the ambient medium during the crossing of the reverse shock. This energy transfer appears after the reverse shock has disappeared and the authors suggest that the transfer is delayed with respect to the traditional case (without Poynting flux). There are thus at least two reasons to have a delayed energy transfer to the forward shock. However, it is unclear how long the delay would be and whether one of these mechanisms can account for the observed shallow decays. As noted by Zhang (2007a), detailed numerical simulations are still missing to check these points. But it is not unrealistic to attribute the shallow decay phase to a slow energy transfer from the ejecta to the ambient medium.

Within the three different scenarios discussed above, it is also necessary to analyze the sudden cessation of the injection at the end of the shallow phase. In the long-lived central engine model this time corresponds to the end of the injection process. In the varying Lorentz factor scenario this time corresponds to a cut-off of the Lorentz factor distribution at the low end, below which the distribution index $s$ is flatter than 1 (so that they are energetically unimportant). In the Poynting flux injection model the well-defined temporal cut-off corresponds to the epoch when all the Poynting energy is transferred to the blastwave (Zhang & Kobayashi 2005). While we have concentrated the discussions on scenarios invoking the late injection of energy, which are the favorites of various authors, other models have been proposed to explain the shallow decay phase. They are discussed now.

### 7.2.2 Different structural jet models

The idea of structured jet models is to consider a jet with distinct spatial components, some of which only contribute to the afterglow emission at later times. One possibility arises if the radiation emitted by components that do not move along our line of sight is strongly beamed away from us at early times and becomes visible later when deceleration causes their beaming cones to widen. This situation might be realized in the ‘patchy shell’ model (Kumar & Piran 2000) as well as in the off-beam jet model (Eichler & Granot 2006). Another possibility exists if some of the ejecta have lower initial Lorentz factors that result in longer deceleration times; early on only a small fraction of the injected energy is transferred to the shocked external medium. This
case is quite similar to case (2) discussed above. It is illustrated by the two-component jet model considered by Peng, Königl, and Granot (2005) as a way of alleviating the pre-Swift constraints, and used by Granot, Königl, and Piran (2006) to interpret the X-ray afterglow light-curves of GRBs displaying the shallow phase. We will now consider these two models: the off-beam model and the two-component jet model.

**Off-beam jet model**

Eichler and Granot (2006) proposed that the γ-rays and the afterglow are not due to the same components of the outflow or more generally that the ratio of γ-ray to baryonic energy per solid angle may vary considerably within the outflow. The consequences of what they call a GRB/baryon anisotropy is the presence of lines of sight that are more favorable for GRB detection whereas others are more favorable for the detection of the early afterglow. Flat early afterglows are produced for viewing angles slightly outside the region that produces bright afterglow emission. These off-beam viewing angles are offset from the prominent early afterglow emission but not from the prompt γ-ray beam. The regions of prominent afterglow emission and of prominent γ-ray emission do not coincide. Under these conditions the shallow phase would be a combination of the decaying tail of the prompt emission and the delayed onset of the afterglow emission seen from viewing angles slightly outside the edge of the jet. This is equivalent to saying that the afterglow onset might be delayed for an ‘offset viewer’ (Panaitescu & Mészáros 1999, Moverski, Sikora, & Bulik 2000, Granot et al. 2002, Dado, Dar, & De Rújula 2002, Dalal, Griest, & Pruet 2002). Eichler and Granot (2006) have shown that the shallow decay phase of the afterglow of GRB 050315 can be explained with a Gaussian jet seen off-axis (Figure 7.3). The light-curves obtained for such Gaussian jets and for various observer offsets \( \theta_{\text{offset}} \) relative to the jet axis are given in Figure 7.4 (from Panaitescu 2006). As \( \theta_{\text{offset}} \) increases, the shallow phase emerges at later time \( (t_{\text{s2}}) \) and at lower flux \( (L_{\text{S2}}) \), and it exhibits a flatter decay \( (\alpha_{\text{s2}}) \). Panaitescu (2006) indicated that this type of structured outflow can be tested by searching for correlations between \( L_{\text{s2}} \) and \( \alpha_{\text{s2}} \) and anti-correlations between \( t_{\text{s2}} \) and \( \alpha_{\text{s2}} \). Panaitescu (2007a) studied 32 bursts with well-defined shallow phases and showed that they did not confirm the correlations expected for this model. Eichler and Granot (2006) noted that with this model, the usual measurements of the γ-ray efficiency based on the comparison of \( E_{\text{iso}} \) with the kinetic energy of the afterglow, do not represent the true γ-ray efficiency. This off-beam jet model also was discussed by Toma et al. (2006) within the framework of the patchy jet model. A very inhomogeneous jet is made with multiple subjets. The superposition of the afterglows of several off-axis subjets produces the shallow decay phase. This phase arises prior to the merging of the subjets. The whole jet produces the conventional late afterglow after the subjets have merged. The shape of the early light-curve depends sensitively on the assumed sideways expansion of the subjets, which is quite uncertain. Depending on this value the shallow decay phase can disappear as discussed by Toma et al. (2006). Their model predicts that events with dim prompt emission will have afterglow light-curves without shallow decay.
Two-component jet model

Granot, Königl, and Piran (2006) used this model to analyze the early X-ray afterglow light-curves discovered by Swift. The model includes a narrow and initially highly relativistic (Lorentz factors greater than 100) conical outflow, which is the source of
the prompt emission, and a mildly relativistic flow (Lorentz factor $\sim 10$) inside a wider coaxial cone that decelerates at later time. It is further assumed that the line of sight of the observer lies within or very close to the opening angle of the narrow component. The broad component becomes visible after it has decelerated in the circumburst medium, and it only makes a contribution to the afterglow emission. In the context of this model the shallow phase corresponds to the emergence of the wide component around its deceleration time. A sketch of this two-component jet model is given in the upper panel of Figure 7.5 (from Panaitescu 2006). The lower panel of Figure 7.5 (Granot, Königl, & Piran 2006) shows the application of this model to the X-ray light-curve of GRB 050315 and the effect of varying the model parameters. A good fit can be obtained with plausible ranges of kinetic energies, opening angles of the two components, and of the Lorentz factor of the dominant wide component. Granot, Königl, and Piran (2006) showed that this fit leads to a kinetic energy of the wide component which is dominant with respect to the narrow one (by a large factor, $\sim 10^2$; see Figure 7.5(b)). The authors propose that the narrow component corresponds to a baryon-poor black-hole driven outflow (Levinson & Eichler 2003), while the wide component could be a baryon-rich disk-driven outflow (Vlahakis & Königl 2003). Jin et al. (2007) used this model to fit the multi-wavelength afterglow of the short GRB 051221A. The quality of the fit is good but they have included additional free parameters in their model. Moreover, the authors indicated that other models can account for the afterglow data, such as a GRB outflow with a range of bulk Lorentz factors (Rees & Mészáros 1998).

### 7.2.3 Prior activity

To avoid the problem of a very high $\gamma$-ray efficiency of the prompt GRB, Ioka et al. (2006a) proposed two models: the first invoking prior activity of the source and the second invoking time-dependent microphysical parameters. The first model introduces a small explosion, with less energy than the main burst that occurs $10^3$ to $10^6$ s before the main burst. Taking into account this prior activity, they show that the forward shock emission due to the main burst ejecta has a shallow decay ($t^{-1/2}$) when the main fireball sweeps the remnants of the precursor that have a low density and a high velocity, and a conventional decay ($t^{-1}$) after that. Since the prior explosion has less energy than the prompt burst the efficiency crisis is avoided.

### 7.2.4 Varying microphysical parameters

The evolution of microphysical parameters has been widely considered (Ioka et al. 2006a, Fan & Piran 2006, Granot et al. 2006, Panaitescu et al. 2006a). In the pre-Swift era the observations of late afterglows were consistent with $\varepsilon_e$ and $\varepsilon_B$ staying constant within each GRB, but with a spread from one burst to another which could not be due to model uncertainties, particularly for $\varepsilon_B$ which shows variations by two orders of magnitude, as shown by Yost et al. (2003). These authors did not find assumptions that would allow fitting the data with universal microphysical parameters. With the exploration of the early GRB afterglows permitted by Swift it is interesting to
consider whether a variation of the microphysical parameters can explain the new numerous early afterglow observations. This is not unreasonable because the energy transfer from protons to electrons and magnetic fields in relativistic shocks is not well understood, as was noted in Chapter 5.
The electron energy fraction, in particular, might vary in the early afterglow (Ioka 2005). Taking into account the fast-cooling electrons emitting the X-ray component and the weak dependency of $L_x$ on the magnetic energy fraction $L_x \sim \varepsilon_e^{(p-2)/4}$ (with $p \sim 2$ for the index of the power-law electron distribution), the connection between the X-ray luminosity $L_x$ and the total luminosity (dissipated power) $L$ can be written as: $L_x = \varepsilon_e L$. As $L \propto t^{-1}$ and $L_x \propto t^{-1/2}$ during the shallow X-ray light-curve, Ioka et al. (2006a) suggested that $\varepsilon_e \propto t^{1/2}$ with a saturation at the equipartition value when the shallow phase ends. Of course the shallow phase is correctly reproduced, with the $-1/2$ slope of the X-ray light-curve. The authors evaluate the consequences of such a model. As an example, the possible increase with time of the afterglow radiation efficiency during the early phase of the X-ray light-curve, due to an increase with time of $\varepsilon_e$ and/or $\varepsilon_B$, would be very useful to reduce the estimated $\gamma$-ray radiative efficiency and to alleviate the constraints on the internal shock model (see Section 7.3).

The possibility of avoiding energy injection after the prompt phase is well analyzed in Granot, Königl, & Piran (2006). Using the relation (see Section 7.2),

$$\varepsilon_X(t)E_{K,iso}(t) \propto t F_X(t)$$

it is clear that the shallow phase requires $\varepsilon_X(t)E_{K,iso}(t)$ to increase after the prompt emission. This is why energy injection has been naturally invoked to flatten the slope of the early afterglow. Another way would be to keep $E_{K,iso}$ constant and to increase $\varepsilon_X(t)$. Granot, Königl, and Piran (2006) develop this point. As the efficiency depends on $\varepsilon_e$ and $\varepsilon_B$, increasing one (or both) of these parameters with time leads $\varepsilon_X(t)$ to increase with time. This is the same idea that was proposed by Ioka et al. (2006a): the increase of $\varepsilon_e$ with time. Another possibility to have $\varepsilon_X(t)$ increasing with time is to consider $p < 2$ while keeping $\varepsilon_e$ and $\varepsilon_B$ constant, but this solution also has its difficulties (Granot, Königl, & Piran 2006).

This idea of having the microphysical parameters increasing with time was also proposed by Fan and Piran (2006), and by Panaitescu et al. (2006a). While this hypothesis successfully explains the early X-ray light-curves, it does not explain why the temporal evolution of microphysical parameters effectively terminates at the end of the shallow phase, and it does not provide a physical reason for $\varepsilon_B$ to increase during this shallow phase. This requires the study of the temporal evolution of shock-generated magnetic fields over a long time. The increase of $\varepsilon_e$ with time is also an open question, even if it cannot be excluded that $\varepsilon_e$ may be linked to the variation of the shock parameters. Fan and Piran (2006) have tested four models involving various physical assumptions: energy injection; evolving shock parameters; a model in which the shock energy is given to a small fraction of the electrons rather than to their totality; and a model with a very low varying density of the surrounding medium. Their analysis of these last two models leads them to conclude that they are ruled out by the X-ray data. The two remaining models, energy injection and evolving shock microphysical parameters, are consistent with X-ray afterglow observations, as we have seen, but Fan and Piran (2006) found that they fail to self-consistently reproduce the X-ray and optical afterglows of GRB 050319 and GRB 050401. Of course this is not the final word on this subject. Given these difficulties, Fan and Piran
(2006) looked for alternatives, considering for instance that the low fluxes of the X-ray afterglow after the prompt emission can be considered as a sign of a ‘missing energy problem’. The flat X-ray emission during the shallow phase might be considered as a case of missing X-ray emission where the energy is dissipated in a different channel outside the X-ray energy range. A tentative attempt to introduce inverse Compton cooling did not work, leading Fan and Piran (2006) to suggest that it would be interesting to look for unexplored processes. Before leaving the explanations of the shallow phase, we now evoke other possibilities which are perhaps less conventional.

7.2.5 Other models

Shao and Dai (2007) suggested that small-angle scattering of X-rays by dust surrounding the source (an X-ray halo) could also give rise to a shallow decay phase. Dado, Dar, and De Rújula (2006) explained early afterglow light-curves with the cannonball model. They mention that the early X-ray afterglows are precisely what is predicted by this model, predictions already obtained before the launch of Swift (Dado, Dar, & De Rújula 2002). Their Figure 2 which compares the cannonball and standard models using XRT data from O’Brien et al. (2006) is quite interesting, leading the authors to a scathing questioning of the fireball model.

For some authors, however, the slow decay phase implies a more drastic revision of the standard model. We now discuss new and non-conventional ideas independently proposed by Uhm and Beloborodov (2007) and by Genet, Daigne, and Mochkovitch (2007).

7.2.6 New ideas: the role of the reverse shock

The idea of these authors is to attribute the early afterglow to the sole processes taking place within the ejecta (internal and reverse shocks) with no contribution from the forward shock, which is assumed to be radiatively inefficient. This demands ejecta with a large range of Lorentz factors, extending down to the lowest values. This idea was proposed to avoid some of the difficulties faced by the standard model after the early X-ray afterglow observations of Swift (see below).

Genet, Daigne, and Mochkovitch (2007) note that when the ejecta has a broad range of Lorentz factors the reverse shock remains active for long times. They further note that the temporal evolution of the power dissipated in the reverse shock is strikingly similar to the temporal evolution of the early X-ray afterglow. They discuss the conditions required for a fraction of the power dissipated in the reverse shock to be radiated into the X-ray range. For that to occur, two conditions must be met: the Lorentz factor of the material ejected during the last stages of source activity must decrease to small values $\Gamma < 10$. The second condition is that a large part of the shock energy must be transferred to only a small fraction ($\xi < 10^{-2}$) of the electrons. Let us analyze the first condition; Genet, Daigne, and Mochkovitch (2007) consider ejecta with high Lorentz factor ($\Gamma = 200$) during 5 s, progressively slowing down to small $\Gamma$ (of the order of unity) until time $t_m = 10$ s, at which time the central engine stops...
ejecting relativistic material (see their Figure 1). The main and first condition of this model is that the Lorentz factor of the material ejected at late times by the source has to decrease to a small value, $\Gamma_f < 10$. With this distribution of Lorentz factors the reverse shock is present for a much longer time. In their simplified model, they calculate the power dissipated in the reverse shock for different environments (uniform ISM and wind; see their Figure 2). The case of a wind is particularly striking; a plateau is observed at about 100 s, similar to what is observed by Swift in the early X-ray afterglow. However, this figure gives the power dissipated in the reverse shock, which is most often expected to appear in the visible/NIR domain (e.g. Sari & Piran 1999). For instance the reverse shock was proposed to explain the early optical flashes associated with several GRBs (see Chapter 6). In order to radiate in X-rays, $\gamma_e$, the typical electron Lorentz factor in the frame of the shocked material must be large, since the emission is due to synchrotron radiation. We have seen that $\gamma_e \propto e(\varepsilon_e m_p)/(\xi m_e)$ (where $e c^2$ is the energy dissipated per unit mass in the comoving frame, $m_p$ and $m_e$ are respectively the proton and electron masses, $\varepsilon_e$ is the fraction of dissipated energy transferred to the electrons and $\xi$ is the fraction of the electrons which receives the fraction $\varepsilon_e$ of the dissipated energy). Large values of $\gamma_e$ can be obtained with small values of $\xi$ (for constant $\varepsilon_e$). If $\xi = 10^{-2}$, as the authors propose, $\gamma_e$ increases by a factor of 100, $E_s$, the characteristic synchrotron energy increases by $10^4$ ($E_s \propto \gamma_e^2$), and $t_s$, the synchrotron characteristic cooling time, decreases by a factor of 100 ($t_s \propto \gamma_e^{-1}$), making it possible to reach the fast-cooling regime ($E_s$ and $t_s$ are given in the rest-frame of the shocked material). With such low values of $\xi$ the emission can be in the X-ray range; this is the second condition given by the authors. This condition indicates that a large part of the reverse-shock energy is transferred to only a small fraction of the electron population. Figure 7.6 shows modeled X-ray light-curves in the 0.3–10 keV range, obtained with an equipartition of the reverse-shock energy ($\varepsilon_e = \varepsilon_B = 1/3$) and $\xi = 10^{-2}$. These light-curves, especially for the wind model, are very similar to those of the early X-ray afterglows observed by Swift, with a clear shallow phase. Since the optical emission from the reverse shock is not completely suppressed, chromatic breaks are expected when $\nu_m$ crosses $\nu_{vis}$. The authors propose that in some bursts the entire afterglow could be due to the reverse shock.

The same conclusion was also reached by Uhm and Beloborodov (2007), with a forward shock which is invisible while the afterglow is due to a long-lived reverse shock in the ejecta. The central engine ejects a relativistic flow which can be considered as a sequence of shells, each shell being characterized by its energy and Lorentz factor $\Gamma_{ej}(\tau)$, where $\tau$ is the ejection time. The ejecta pushes the ‘blast’ (a thin shell of compressed gas between the forward and the reverse shock). The ‘blast’ is viewed as one hot body which moves with a common $\Gamma$ but it is made up of different hot shells that pile up from the forward and reverse shock (Beloborodov & Uhm 2006). The standard model is used to calculate the synchrotron emission from the shocked medium (forward-shock emission) and shocked ejecta (reverse-shock emission). Figure 7.7 gives the stratification function $\Gamma_{ej}(E_{ej})$, where $E_{ej}(\tau)$ is the kinetic energy of the ejected flow ($E_{ej} = 0$ corresponds to the first ejected shell, $E_{ej} = E_b$ corresponds to the last shell, $E_b$ is the total energy of the explosion). The reverse
shock crosses the head at early time, \( t_{\text{obs}} < 10^2 \text{ s} \), and proceeds to the tail. Figure 7.8 summarizes the main results of the numerical model. From this figure it appears that the X-ray and optical emission produced by the forward shock (FS) are inconsistent with observations, so Uhlm and Beloborodov (2007) propose that the FS is invisible due to the low amplification of the magnetic field or to the inefficient acceleration of the electrons. On the other hand X-ray afterglows due to the reverse shock are much weaker at \( 10^2 \)–\( 10^4 \) s but they present a very steep decay followed by a shallow decay as observed in early X-ray afterglows. The steep decay appears when the reverse shock enters the ‘tail’ where the ejecta density drops dramatically because of the large

Figure 7.6. Calculated X-ray light-curves in the 0.3–10 keV range for various reverse shock models:
(a) A wind model with different values of \( A_e \) (\( A_e = 1 \) for a wind mass loss rate \( M_w = 10^{-5} \, M_\odot \text{ yr}^{-1} \)). Three values of \( A_e \) are considered: 0.5, 0.1, and 0.05 plotted with dashed, full, and dotted lines respectively.
(b) A uniform density model with \( n = 1000 \) (dashed line), 10 (full line), and 0.1 cm\(^{-3} \) (dotted line). In both panels it is assumed that \( \varepsilon_e = \varepsilon_B = 1/3 \) and \( \xi = 10^{-2} \) (see text). The redshift is assumed to be \( z = 2 \) (from Genet, Daigne, & Mochkovitch 2007).
Figure 7.7. Stratification function $\Gamma_{ej}(E_{ej})$ in the model of Uhm and Beloborodov (2007, see text). $E_{ej} = 0$ corresponds to the first ejected shell and $E_{ej} = E_h$ corresponds to the last shell. The reverse shock starts at $E_{ej} = 0$ and moves toward $E_{ej} = E_h$ passing through points A, B, C, D. The transition from head to tail of the ejecta occurs between points A and B (Uhm & Beloborodov 2007).

The shallow phase [Sec. 7.2]

The shallow phase begins at $t_{obs} \sim 10^3$ s, when the reverse shock propagates into the region corresponding to the flatter stratification function (smaller $d\Gamma_{ej}/dE_{ej}$ and higher $\rho_{ej}$, see Figure 7.7). The final break at point D occurs around $t_{obs} \sim 10^5$ s, and it corresponds to the reverse shock reaching the end of the ejecta. The authors also discussed the optical light-curves, which are found to exhibit a shoulder accompanying the X-ray plateau. The authors studied the effect of bumps in the external density; they showed that these bumps have a much stronger impact on the optical and X-ray light-curves of the afterglow than in standard forward-shock models. In the discussion the authors analyzed the major problems of the forward-shock model and showed how the reverse shock is able to solve them. Among their five points, they discussed the critical question of the shallow phase (and the need for energy injection in the forward-shock model) and concluded that the reverse shock model can explain it naturally. They showed that the plateau BC in the X-ray afterglow (Figure 7.8(e)) is produced when the reverse shock recovers after the steep transition to the tail of the ejecta and enters the shallow part of the stratification function $\Gamma_{ej}(E_{ej})$ (Figure 7.7). In their scenario $E_{tail} \sim E_{head}$, and a relatively low efficiency of the prompt emission is needed (1–10%). Another point they analyzed is the steepening at point C after the plateau (Figure 7.8(e)); in the forward-shock interpretation, this is generally considered as the end of energy injection and the steepening is expected to be present in all the energy bands. But, as indicated by Panaitescu et al. (2006a), the break is often present only in the X-ray domain, so this is also an open issue. Uhm and Beloborodov (2007) showed that their model predicts no abrupt changes in $\beta_x$, the spectral index in X-rays, and only a shoulder (and not a break) in the optical light-curve (Figure 7.8(e) and (f)). The second break, interpreted as a jet break in the standard model, is also explained with the reverse shock: it appears when the shock reaches the end of the ejecta or faces a steep decline in $\rho_{ej}$.

In these two models, the afterglow is produced by the shocked ejecta and not by the shocked external medium. The ejecta is made of two parts: the head and the tail, with different $\Gamma_{ej}$ and comparable kinetic energies. The afterglow light-curves are strongly shaped by the stratification of the ejecta. A large variety of afterglows can be obtained by changing the stratification function $\Gamma_{ej}(E_{ej})$; for instance, the plateau disappears if $\Gamma_{ej}(E_{ej})$ decreases smoothly after point A (Figure 7.7). Finally, Uhm and
Figure 7.8. Blastwave evolution in the reverse shock (RS) model of Uhm and Beloborodov (2007): (a) Density of the pre-shock ejecta $n_{\text{ej}}(\text{RS}) = \rho_{\text{ej}}/m_p$ as a function of $r$, the radius of the expanding blastwave. The density of the external medium is shown for comparison, $n_1 = 1 \text{ cm}^{-3}$. (b) Lorentz factor of the pre-shock ejecta, $\Gamma_{\text{ej}}(\text{RS})$, and of the blast, $\Gamma$, as a function of $r$. (c) Pressure at the forward and reverse shock $p_f$ and $p_r$. (d) Relative Lorentz factor of the RS. (e) R-band and soft X-ray light-curves of the afterglow due to the forward and reverse shocks ($\varepsilon_e = 0.1, \varepsilon_B = 0.01$, and $p = 2.3$). Thin curves give the forward shock (FS) emission and thick curves give the RS shock emission with the plateau observed in the X-ray band, and a shoulder in the optical band. (f) Evolution of spectral indices $\beta_e$ (X-rays) and $\beta_0$ (optical) of the RS emission.

Beloborodov (2007) predicted that the reverse shock must be very bright in the infrared bands. This is a good way to check the model.

We have gone into the details of these two ideas because they are quite unconventional and they could solve many difficulties linked to the standard forward-shock model. In these models the afterglow may carry important information about the explosion since it is produced by the shocked ejecta. Nevertheless, as for every new idea, numerous theoretical and observational verifications, using multi-wavelength afterglow observations, are needed. It is too early to abandon the well studied and checked (for many aspects) standard forward shock model as the explanation for the GRB afterglows.

The discovery of the early X-ray afterglow, its properties, and the possible need for substantial energy injection after the prompt phase exacerbate the question of the gamma-ray efficiency of the internal shocks (see Section 5.3.3 for a discussion on this subject in the pre-Swift period). Let us now go to a new analysis of this crucial topic after the discovery of the early X-ray afterglows.

### 7.3 VISITING THE EFFICIENCY OF INTERNAL SHOCKS

The problem of the radiative efficiency of internal shocks is quite well analyzed by Granot, Königl, and Piran (2006), so we will use the main points of their analysis.
Let’s call $\varepsilon_\gamma$ the $\gamma$-ray efficiency (i.e., the fraction of the energy injected at the source which is converted into the prompt radiation). This efficiency can be connected with $E_{\text{iso}}$ and $E_{K,\text{iso}}(0)$ which are respectively the observed isotropic energy in the prompt GRB phase and the isotropic equivalent kinetic energy in the afterglow at $t_{\text{dec}}$, i.e. at the beginning of the afterglow phase.

The relation between these quantities is:

$$\varepsilon_{\text{GRB}}\varepsilon_\gamma/(1 - \varepsilon_\gamma) = E_{\text{iso}}/E_{K,\text{iso}}(0) = kf$$

where $\varepsilon_{\text{GRB}}$ is the fraction of radiated energy that falls within the observed photon energy range.

$$k = E_{\text{iso}}/E_{K,\text{iso}}(t_*) \quad \text{and} \quad f = E_{K,\text{iso}}(t_*)/E_{K,\text{iso}}(0)$$

where $t_*$ is an arbitrary time where it is convenient to measure $E_{K,\text{iso}}$. We have seen in Section 6.7.2 that the X-ray luminosity at $t = 10$ h has been widely used for this purpose. This time is typically beyond the end of the shallow phase. The parameter $k$ contains the information about the radiative efficiency of internal shocks, while $f$ describes the injection of kinetic energy in the ejecta.

If $\varepsilon_{\text{GRB}}$ can be considered to be 1, we see that $\varepsilon_\gamma/(1 - \varepsilon_\gamma) = kf$. The $k$ term connects $E_{\text{iso}}$ and $E_{K,\text{iso}}$ at 10 h, and we have already discussed the value of this component. The factor $f$ is new and it is introduced if there is a significant energy injection (for instance to explain the shallow phase) in the afterglow between time $t = 0$ and $t = 10$ h.

Before Swift, $\varepsilon_\gamma/(1 - \varepsilon_\gamma)$ was equal to $k$ and written $k = (E_{\text{iso}}/E_{K,\text{iso}})$, which is equivalent to $\varepsilon_\gamma = E_{\text{iso}}/(E_{K,\text{iso}} + E_{\text{iso}})$. The pre-Swift analyses led to the measurement of a rough equality between the isotropic equivalent $\gamma$-ray and late (10 h) kinetic energies; under these conditions $\varepsilon_\gamma$ is around 0.5.

For Swift GRBs, Fan and Piran (2006) gave an efficiency $\varepsilon_\gamma \sim 0.5$, even when they took into account the possibility of energy injection. This is due to the fact they found a low value, $\sim 0.1$, for the efficiency of the pre-Swift GRBs. They think that such values do not challenge the standard internal-shock model. Nevertheless, these are high values, which can be considered as fairly restrictive for the internal-shock model. Peng, Königl, and Granot (2005) and Granot, Königl, and Piran (2006) have formulated the conditions which allow the internal shock model to work with such high efficiencies.

Granot, Königl, and Piran (2006) revisited the evaluation of $k$, and more specifically the value of $E_{K,\text{iso}}(10 \text{ h})$ calculated from $F_{\gamma}(10 \text{ h})$ using the standard afterglow theory. For 17 pre-Swift GRBs with known redshifts they observed large differences between the values of $E_{K,\text{iso}}(10 \text{ h})$ calculated by different groups; in particular, many values were significantly larger than the values estimated by Lloyd-Ronning and Zhang (2004). Using values of $E_{K,\text{iso}}$ that are too small results in an overestimate of the $\gamma$-ray efficiency of pre-Swift GRBs (see also Fan & Piran 2006). This is confirmed by late-time radio observations, which provide independent estimates of the GRB kinetic energy. Using radio calorimetry Berger, Kulkarni, and Frail (2004) found high values of $E_{K,\text{iso}}$ and correspondingly low $\gamma$-ray efficiency. The values of $E_{K,\text{iso}}$ calculated by Granot, Königl, and Piran (2006) lead to typical values $k = 0.1$–0.2, using their fiducial microphysical parameters $\varepsilon_e = 0.1$, $\varepsilon_B = 0.01$, and
$p = 2.2$. These values are similar to those calculated by Fan and Piran (2006). Microphysical parameters play an important role in the estimate of $E_{K,iso}$; using microphysical parameters derived from the multi-wavelength modeling of the afterglows (Panaitescu & Kumar 2002) the values of $E_{K,iso}(10 \text{ h})$ are significantly lower, resulting in higher values of $k$, close to unity. Finally, the values of $E_{K,iso}(10 \text{ h})$ have a large dispersion (see table 1 from Granot, Königl, & Piran 2006) so the question of the efficiency crisis of the internal shocks must be considered with caution.

Now we can look at the new results of Swift, following the formalism of Granot, Königl, & Piran (2006). With the early shallow phase, the factor $f$ is back and $\varepsilon_\gamma$ can be estimated from: $\varepsilon_\gamma/(1 - \varepsilon_\gamma) = kf$, assuming $\varepsilon_{\text{GRB}} = 1$.

If the shallow decline is due to energy injection, the typical value of $f$ is $f \sim 10$, indicating that the initial afterglow energy $E_{K,iso}(0)$ is significantly smaller than the value of $E_k$ determined at 10 h (Granot, Königl, & Piran 2006). So for a value of $k \sim 0.1$, as reported using the fiducial parameters, we have again $\varepsilon_\gamma/(1 - \varepsilon_\gamma) \sim 1$ and a high efficiency $\varepsilon_\gamma \sim 0.5$. This is also the conclusion reached by Fan and Piran (2006). Of course the choice of the fiducial microphysical parameters, which affects significantly the values of $k$ and $E_{K,iso}(10 \text{ h})$, is somewhat arbitrary. If, on the other hand, we consider pre-Swift values with $k \sim 1$ and keep $f \sim 10$, we are confronted with a true efficiency crisis because $\varepsilon_\gamma$ reaches 0.9. These are the values estimated by Ioka et al. (2006a) for two models with energy injection.

Various solutions have been proposed to avoid the need for extremely radiatively efficient internal shocks. First, the efficiency crisis can be avoided if $k \sim 0.1$ and if the shallow decline does not arise from energy injection, i.e. $f \sim 1$. In this case $\varepsilon_\gamma \sim 0.1$ and the internal shock model is acceptable. Ioka et al. (2006a), for instance, have proposed solutions to explain the shallow phase without energy injection (see Sections 7.2.3 and 7.2.4). Second, since the high values of $k$ inferred from pre-Swift afterglow fits are obtained on the basis of the assumptions of standard afterglow theory, it is always possible to change some of these assumptions. For instance, if only a fraction $\xi_e < 1$ of the electrons is accelerated to relativistic energies, Eichler and Waxman (2005) show that $E_{K,iso}$ is increased by a factor $\xi_e^{-1}$ while $k$ is reduced by this same factor. This is a good way to have a lower value of $k$ and to reduce the efficiency problem. Another solution suggested by Granot, Königl, and Piran (2006) is the possibility of having a fraction $\varepsilon_\gamma$ of the total energy ending up in a totally different form: for instance, in low-frequency electromagnetic waves or into high-energy cosmic-rays and neutrinos. This energy does not contribute to the kinetic energy of the afterglow. Finally, $\varepsilon_\gamma$ is multiplied by $(1 - \varepsilon_\gamma)$ and can be significantly reduced if $\varepsilon_\gamma$ is large enough.

Without being exhaustive, but just to illustrate the uncertainties around this question, it has been suggested that at least some GRBs are strongly magnetized or even Poynting-flux dominated, with a direct transfer of Poynting flux into non-thermal radiation (Usov 1994, Thompson 1994, Zhang 2005, Zhang & Kobayashi 2005, Granot, Königl, & Piran 2006). Today it is still difficult to assess the efficiency of this process, which could nevertheless be much more efficient than internal shocks.
The analysis of Granot, Königl, and Piran (2006) (see also the conclusions of Zhang et al. 2007) illustrates the large uncertainties existing on the measurement of $k$, complicated by the uncertainty about the explanation of the shallow phase, with or without energy injection, and about the measurement of $f'$. In view of these uncertainties, the internal shock model is still alive but it has a sword of Damocles above its head with the clues pointing towards a very high radiative efficiency of the prompt GRB, which is possibly too high for internal shocks. As we have seen, this puzzling situation has already led Genet, Daigne, and Mochkovitch (2007) and Uhm and Beloborodov (2007) to introduce dramatic changes in the explanation of the early afterglow (or even of the full afterglow), using only the reverse shock with an inefficient forward shock.

We have seen that the early X-ray afterglows of many GRBs are characterized by a steep decay phase followed by a shallow phase. This general behavior was not predicted by standard afterglow models and challenges them. Swift made two other important discoveries: the presence of X-ray flares and of chromatic X-ray breaks not simultaneously seen in the optical domain. Let us turn first to the discovery of the early X-ray flares and the analysis of their origin.

### 7.4 X-RAY FLARES

Before analyzing the possible mechanisms that can explain the X-ray flares, it is useful to recall here their observational properties, which have been explained in detail in Chapter 4 and in Burrows et al. (2005) and Chincarini et al. (2007). These flares have a rapid rise and fall; many of them seem to be superimposed on a smoothly decaying afterglow component; multiple flares are observed in some bursts; the flux increase is usually large and in a few cases (e.g. GRB 050502B; see Figure 4.14) the fluence of the flare can be comparable with that of the prompt emission; flares soften as they progress and later flares are less energetic and broader than early flares.

A very complete review of the mechanisms that have been invoked to produce these X-ray flares is given by Zhang et al. (2006). We will use it extensively to discuss those mechanisms. We will first discuss various mechanisms which have been studied but found to be unable to account for the majority of the X-ray flares, and then the models involving a reactivation of the central engine, which seem to be able to reproduce the observed properties of the X-ray flares. These models can work if the activity of the central engine extends as late as 1 day after the trigger. Here again, we see a new behavior, which is essentially a surprise and requires adjustments to the pre-Swift models.

#### 7.4.1 Models which have difficulties

**Emission from the reverse shock**

While it is in principle possible that synchrotron (or synchrotron self-Compton) emission from the reverse shock dominates that of the forward-shock emission in
the X-ray band, the lack of strong UV–optical flares in the UVOT observations led Zhang et al. (2006) to rule out a reverse-shock origin for the X-ray flares. Furthermore, this model could explain a single X-ray flare but it cannot be invoked when several flares are observed. Kobayashi et al. (2007) considered synchrotron self-Compton radiation from the reverse shock when the inverse Compton process dominates the cooling of the shocked electrons. They showed that for plausible parameters this scattering process can produce an X-ray flare with a relative amplitude change of a factor of several. However, flares with larger amplitude and multiple X-ray flares in a single GRB are likely to be produced by another mechanism.

**High-density clouds surrounding the progenitor**

Clouds surrounding the GRB progenitor may exist. These density clumps have been invoked by Dermer & Mitman (2004) to interpret the variability of the prompt GRB emission. Zhang et al. (2006) calculated the properties of an X-ray flare powered by a relativistic fireball interacting with a dense cloud. They showed that the rising phase could be very sharp but the decaying slope stays rather flat and does not reproduce the steep decay of typical X-ray flares (their Figure 4). Detailed calculations suggest that the variations caused by density inhomogeneities are generally not very significant (Zhang et al. 2006, Nakar & Granot 2007; but see Ramirez-Ruiz et al. 2005). Ioka, Khobayashi, and Zhang (2005) determined limits on the timescale and amplitude of variabilities in GRB afterglow and used them to identify the possible sources of afterglow variabilities. In particular they give a limit on $\Delta F_\nu/F_\nu$ (the relative variability flux) for bumps due to density fluctuations, and they show that X-ray flares violate these limits.

**Patchy jets**

The existence of jets with large energy fluctuations has been proposed by Kumar and Piran (2000); this has been called the patchy jet model. This model can be considered as a variation of the two-component jet model and it has the same problems producing X-ray flares. After each flare, the afterglow level is boosted and would not return to the previous level as observed after the X-ray flares. Also the model cannot explain the fast decays of X-ray flares.

**Late energy injection**

In the internal-shock model collisions between shells explain the prompt GRB emission. Zhang and Mészáros (2002) studied the late injection of high-$\Gamma$ shells catching up with the decelerating blastwave, leading to the re-energization of the forward shock. Such collisions would give bumps in the light-curve due to the increase of the total energy in the blastwave. This energy injection might give a flare but after the flare the flux level does not return to its previous level, and here again the decay slope is inconsistent with the steep decay of the X-ray flares.
Neutron signature

While it is most often assumed that the ejecta consists of protons, electron/positron pairs, and radiation, Derishev, Kocharovsky, and Kocharovsky (1999a,b) investigated the dynamics and the possible observational signatures of a relativistic neutron-rich fireball, and Beloborodov (2003) studied the nuclear composition of GRB fireballs. Among free nucleons, α-particles and deuterium, there is the survival of a significant neutron component and a possible neutron excess characterized by a neutron-to-proton ratio above unity. The implications of this neutron component have led various authors (Derishev, Kocharovsky, & Kocharovsky 1999a, Baheall & Mészáros 2000, Mészáros & Rees 2000, Waxman 2000, Waxman & Baheall 2000) to consider the fireball as an interesting source of multi-GeV neutrinos (see Chapter 9). A second implication analyzed by Beloborodov (2003) concerns the dramatic impact of the neutron component on the explosion dynamics at radii as large as $10^{17}$ cm. The neutron component changes the deceleration of the fireball by the external medium because neutrons fully decouple from ions at late stages of the explosion and coast with a constant Lorentz factor. As the ion fireball decelerates, the neutrons form a leading edge. But this edge decays due to the mean lifetime of the neutrons, leaving behind a trail of decay products mixed with the ambient medium. The ion fireball sweeps up the neutron decay products, driving a shock wave into it. The afterglow emission is produced in the neutron trail, and a rebrightening (a bump in the light-curve) and a spectral transition are expected at $R = R_{\text{trail}} = 10^{17}$ cm when the impact of the neutrons turns off (Beloborodov 2003).

Vlahakis, Peng, and Königl (2003) also reported the possible presence of bumps due to neutrons in the afterglow light-curves. They considered disk-fed, neutron-rich, outflows ejected by a black hole, with a neutron-to-proton ratio which might be as high as 30. Their results were obtained with an exact semi-analytic relativistic MHD solution. Fan and Wei (2004) predicted the presence of detectable long-wavelength flashes during the prompt $\gamma$-ray emission phase, when there is a large amount of neutrons within the GRB outflow. In their neutron-rich internal shock model the Lorentz factors of the proton shells and of the accompanying neutron shells are variable. If the GRB lasts long enough, the slow neutron shells ejected at earlier times will be swept up by the faster proton shells ejected at later times. As the slow neutron shells decay continuously the proton shells interact with the $\beta$-decay products of the slow neutron shells and then may give rise to detectable long-wavelength prompt emission.

Fan, Zhang, and Wei (2005a,b) calculated the early optical afterglow light-curves of neutron-fed GRB fireballs for various fractions of neutrons and circum-burst media, including the reverse shock. First, the neutron component is inevitable for a baryonic fireball, and depending on the baryonic load the effect of the neutron component is variable. As a consequence, the neutron signature should be absent if the fireball is strongly dominated by a Poynting flux and has extremely low baryon loading. They used their model to explore the very early optical/NIR emission from GRB 041219A, assuming that the ejecta are neutron-rich and that the optical emission is powered by the proton shells interacting with the neutron decay products at a
distance much greater than the typical internal shock radius. Overall, the presence of free neutrons in the outflow is equivalent to some late injection of energy in the afterglow, which cannot explain the X-ray flares, as we have already seen.

To summarize, if the X-ray flares are associated with the external shock, a lot of energy is needed to give a significant signature on top of the afterglow light-curve. As this energy does not disappear after the flare, the afterglow flux level cannot resume its previous level at the end of the flare. This is the fundamental reason which allowed many authors to rule out or strongly disfavor the models related to the forward-external shock (Fan & Wei 2005, Romano et al. 2006, Wu et al. 2006, Zhang et al. 2006, Lazzati & Perna 2007, Chincarini et al. 2007). We now discuss the possibility that X-ray flares are due to internal shocks.

7.4.2 Flares and the late central engine activity

After the prompt emission, which explains the GRB itself, the reactivation of the central engine at later times could produce X-ray flares. In order to avoid the problems of the models based on the forward shock, these flares must be due to internal energy dissipation processes taking place before the deceleration of the ejecta.

Zhang et al. (2006) considered two variants of this model:

(a) The central engine ejects more energy at a late time in the form of an unsteady wind with varying Lorentz factors. These discrete shells collide with each other and produce the observed emission.

(b) The late injection could be mainly in the form of magnetic fields/Poynting flux, and the X-ray flares are due to the intermittent dissipation of the magnetic fields, possibly through magnetic field reconnection.

Lazzati and Perna (2007) showed that late episodes of ‘prompt emission’ can have two different origins in the late internal-shock model. The first assumes that the inner engine itself is active for a time comparable with the X-ray flaring activity. The second considers a central engine which is short-lived, but produces a tail of slower material, in addition to the fast ejecta. This slower material can produce X-ray flares via internal dissipation (and therefore prompt emission) at late times. The slow shells ejected after the fast ejecta will not produce an external shock at the typical external-shock radius since the ambient medium has already been swept up by the fast ejecta. So there are basically two scenarios. In the first one the shells are produced at late times by a long-lived engine, in the second one the shells are produced during the GRB phase but they dissipate their energy at much later times. These two situations have been analyzed by Lazzati and Perna (2007). Their conclusion is that a large fraction (if not all) of X-ray flares must be due to the long-lasting activity of the GRB engine rather than to external shock activity (as we have already said) or to the emission of slow shells immediately after the end of the prompt emission.

Another approach has been proposed by Wu et al. (2006) who studied X-ray flares in the context of a central engine with two periods of activity. This is a
simplification because the presence of multiple X-ray flares may indicate that the central engine may have many periods of activity for a long time (see, e.g., Watson et al. 2006, Cusumano et al. 2006b). If the central engine has two periods of activity they found four basic types of X-ray light-curves. They explained some bursts with X-ray flares by late internal shocks and others by late external shocks. In both cases long activity of the central engine is required. To confirm the internal-shock origin of the X-ray flares, Wu et al. (2006) suggested searching them for the spectro-temporal correlations that are known to exist for the GRB prompt emission. They mentioned the spectral lag relation (e.g. Chen et al. 2005) and the spectral width relation (Fenimore et al. 1995). This would be a good test of the common nature of X-ray flares and of the GRB prompt emission. Zou, Dai, and Xu (2006) proposed explaining both the X-ray flare emission and the optical emission of GRB 050904 with internal shocks. Their model assumes that a slow massive shell propagates in front of many faster shells. The faster shells, with similar masses and energies, catch up with the slower shell ahead at different radii. These collisions lead to forward and reverse shocks. In this case, the gamma-rays and early X-rays are mainly emitted from the reverse shock. Their model also requires ‘superlong’ activity of the central engine. Fan, Zhang, and Proga (2005) found that the flares associated with short GRBs are ejected from the central engine and should be of magnetic origin because the jet powering the late X-ray flare must be launched via magnetic processes rather than via $\nu\bar{\nu}$ annihilations and must carry a dominant ordered magnetic field component. As a result the flares are expected to be linearly polarized so this mechanism can be tested. They suggested that the same mechanism may also apply to X-ray flares following long GRBs (see also Burrows et al. 2005, Piro et al. 2005).

Staff, Ouyed, and Bagchi (2007) proposed interpreting the early X-ray afterglow light-curves within the context of a very different model: the quark nova scenario. As in the models that we have already discussed, the activity of the inner engine is extended, but this model has three stages, which are illustrated in Figure 7.9. They successively invoke a NS with an accreting disk (stage 1) which evolves within hours to days into a quark star with its accretion disk (stage 2). Ejecta are then formed, with internal shocks created from colliding shells in the ejected material, producing the GRB. 2–2000 s later, the quark star may collapse into a BH (stage 3) if it has accreted too much matter to be stable. This last episode causes a reactivation of the internal-shock activity, due to the interaction of the ultra-relativistic jet from the BH with the slower ejecta, previously released by the quark star. This new internal shock activity explains the X-ray flares seen in many X-ray afterglows. Staff, Ouyed, and Bagchi (2007) using this three-stage model explained the canonical X-ray afterglow light-curves with their different phases and the X-ray flares (Figure 7.10).

These various studies show that internal shock models present two advantages to explain X-ray flares. First, restarting the central engine is equivalent to resetting the initial time $t_0$ (see the discussion on the choice of $t_0$ to explain the steep afterglow decay in Section 7.1.2). Under these conditions, the X-ray flares and the underlying X-ray afterglow observed at the same observer time $t$ originate from different physical sites at different times. Second, invoking a long activity of the central engine avoids the implication of large energy budgets. Remember that the supplemental energy
Figure 7.9. The three-stage GRB engine proposed by Staff, Ouyed, and Bagchi (2007). First a NS is formed in the core collapse of a massive star, later the NS collapses into a quark star due to spin down or accretion, and finally the quark star collapses into a black hole. Stages 2 and 3 correspond to the production of the GRB and its afterglow; X-ray flares can be produced in stage 3.

Figure 7.10. Illustration of the mechanism leading to the flattening of the X-ray afterglow light-curve and to X-ray flares in the three-stage model of Staff, Ouyed, and Bagchi (2007). If the quark star accretes too much mass, it will collapse into a BH after some time. The jet emitted in the accretion onto the BH can be massive and have a high Lorentz factor; when this outflow collides with a slow shell previously ejected by the quark star an X-ray flare results. Several flares can be produced through collisions with several low Lorentz factor shells (from Staff, Ouyed, & Bagchi 2007).
needed in the forward shock to produce a significant bump in the afterglow must be comparable to the energy that defines the afterglow emission level. Large energy budgets are thus required in external-shock models (Zhang & Meszáros 2002). For internal-shock models the energy budget is reduced because the flares and the afterglow are two different physical components. This energy budget is further reduced if the flares appear at late times when the afterglow is already low. In fact internal-shock models ask for a small fraction of the prompt emission energy (except for exceptional events like GRB 050502B). This eases the requirements on these models. This late internal dissipation model has been tested by Liang et al. (2006) with the same strategy as for the early steep X-ray afterglow phase. The decay of X-ray flares is assumed to be controlled by the curvature effect after the abrupt cessation of the internal dissipation. After subtracting the underlying forward shock afterglow contribution Liang et al. (2006) searched for the initial $t_0$ time which allows, for each flare, the decay slope to satisfy the requirement of the curvature effect. A very convincing result is obtained: most of the flares have their $t_0$ at the beginning of the flare. It seems that this model is robust. Thus, it appears that X-ray flares may have essentially the same origin as the prompt $\gamma$-ray emission, i.e. they are produced by some internal energy dissipation processes which occur before the ejecta is decelerated. Nevertheless, a crucial issue remains: can the central engine restart after being quiescent for some time, $\sim 10^3$ s for several bursts, and sometimes as much as one day after the prompt emission. Before Swift, the possibility to have a central engine with extended activity was only considered to explain the precursors seen in some GRBs, a few hundred seconds before the main burst.

The question of restarting the GRB central engine has been considered by several authors (Burrows et al. 2005, Zhang et al. 2006, Fan & Wei 2005, Ioka, Kobayashi, and Zhang 2005, Wu et al. 2006, Falcone et al. 2006, Romano et al. 2006, Lazzati & Perna 2007, Panaituscu et al. 2006b, Panaitescu 2007a). The major problem is to understand how the central engine may continue to be active minutes to hours after the GRB. Such long-lasting activity provides important constraints on the properties of the central engine, which depend on the possible mechanisms of its reactivation. A summary of these mechanisms can be found in Zhang (2007a) for instance. There are interesting suggestions which have to be studied in more detail by robust numerical simulations. Let us list the present ideas (Zhang, 2007a).

### 7.4.3 The possible origin of the late central engine activity

**Fragmentation or gravitational instabilities in the massive star core**

GRB 050502B was a very remarkable event with its very intense X-ray flare arriving some 400 s after the beginning of the prompt emission and containing as much energy as the initial GRB (the fluence of the X-ray flare in the 0.2–10 keV energy band exceeded the fluence of the prompt burst (Burrows et al. 2005, Falcone et al. 2006). This burst gave rise to a lot of questions and prompted King et al. (2005) to propose a two-stage collapse, which could explain how the central engine could restart or remain active at late times. First, a compact object is formed, which accretes a stellar
mass on a very short timescale, presumably from a neutron torus. This is the origin of the GRB. In order to produce an X-ray flare with energy comparable to the GRB with a delay $t_d \sim 400$ s, King et al. (2005) suggested the formation of a second stellar mass torus that will be accreted after some delay $t_d$. This mass might be a second compact ‘star’, for instance a self-gravitating neutron lump formed because the collapsing core had enough angular momentum to fragment. Gravitational radiation then drags the lump in, toward the first compact object, and tides smash it into a torus that can be accreted. This second accretion episode can release an energy comparable to the original GRB. The time delay between the burst and the flare reflects the time needed by the orbiting fragments to be dragged in due to the emission of gravitational radiation. In this model, X-ray flares result from the fragmentation of the collapsing stellar core and from the subsequent merging of a significant fragment with the most massive compact object formed in the collapse. This idea of a two-stage collapse was also envisaged by Davies et al. (2002), who considered that the collapse of a rapidly rotating stellar core is likely to lead to fragmentation and the formation of more than one compact object of nuclear density. For a sufficiently complex fragmentation process several ‘small’ neutron clumps can be formed and accreted, leading to several, possibly less intense, X-ray flares. The rapid rotation, which seems to be required for the production of gamma-ray bursts (see Section 8.3 on collapsars), might be essential to explain the fragmentation of the core and new starts of the engine at late times, accompanied with long-lasting activity. The model of Davies et al. (2002) can also explain a possible delay (hours to a day) between the burst and supernova, as has been proposed by Reeves et al. (2002) for GRB 011211.

**Fragmentation or gravitational instabilities in the accretion disk**

The fact that flaring activity exists for short GRBs as well as the long ones led Perna, Armitage, and Zhang (2006) to propose that its origin might have something to do with the common feature of the two likely progenitors (see Chapter 8), a hyper-accreting accretion disk around a stellar black hole of a few solar masses. In these systems the accretion proceeds with rates of one to several solar masses per second. Large-amplitude instabilities occurring in the outer part of the accretion disk might be due, as they suggest, to gravitational instability. Fragments of material would be created within the initial hyper-accreting accretion disk, in particular in the outer parts of the disk, where gravitational instability might lead to large amplitude changes in the inner accretion rate or to the complete fragmentation of the disk. The fragments would be dragged into the BH via viscous effects; new fuel for the central engine would be provided via tidal disruption of these fragments. Therefore the flares would be due to blobs of material that initially circularize at various radii and subsequently evolve viscously. MacFadyen, Woosley, and Heger (2001) have also considered long-lasting central engine activity within the context of the collapsar model. It would be due to the fallback of material onto the central black hole after a stellar collapse. In these models, the episodic energy output is driven by changes in the mass supply and subsequently in the accretion rate. The inner part of the accreting system responds to changes in the accretion flow at larger radii.
Figure 7.11. General structural features of the inner MHD flow at three different accretion stages from left to right. (1) The inner flow during the hyper-accretion phase: a relativistic jet forms and a strong poloidal magnetic field is accumulated at the center. (2) Hyper-accretion cannot be sustained because the mass supply rate from the outer inflow drops with time. When the inflow rate is relatively low the magnetic field accumulated in stage 1 can support the gas against gravity. The inflow almost stops at the distance \( r_m \) (the magnetospheric radius). (3) The accretion restarts when the gas accumulated at the edge of the inflow squashes the magnetosphere. The accretion torus is rebuilt and a relativistic jet is reproduced. The third stage would correspond to the time when X-ray flares are produced (from Proga & Zhang 2006).

**Magnetic barrier around the accretor**

Proga and Zhang (2006) proposed that the rate and time behavior of the accretion and the energy output are influenced by the region in the vicinity of the accretor and by the accretor itself. In their model the role of magnetic barriers is crucial to modulate the accretion rate (Figure 7.11). They suggest that the inner flow during the hyper-accretion phase is modulated by the presence of a strong poloidal magnetic field. Figure 7.11 illustrates three different accretion stages when the inflow exists or is halted. For both merger and collapsar models a decrease of the mass supply rate is expected, especially in the late phase of activity, because the mass density decreases with increasing radius. In addition, as mass is accreted onto the BH, the magnetic flux is accumulating in the vicinity of the BH. This magnetic flux can become dynamically important and affect the inner accretion flow, repeatedly stopping and restarting the energy release. Figure 7.11 (from Proga & Zhang 2006) illustrates the effect of the strong poloidal magnetic field. It can interrupt the accretion from the torus for a short time and the GRB activity may be the result of an accretion flow modulated by the ‘magnetic barrier’ and gravity. This might give a natural mechanism for flaring variability in magnetic models of GRBs (see also Fan, Zhang, & Proga (2005) for short GRBs).

**NS–BH merger**

Flares in short GRBs (see for instance Barthelmy et al. 2005) pose problems for merger models. Numerical simulations of NS–NS mergers give a short central engine timescale (0.01–0.1 s; Aloy, Janka, & Müller 2005). The viscous timescales that an
accretion disk can plausibly provide seem difficult to reconcile with the long time-scales of X-ray activity (hours). Nevertheless, extended accretion over several seconds seems possible. Faber et al. (2006), for instance, used relativistic BH–NS merger calculations to see if such events may trigger short hard GRBs; they estimated that about 0.03 \( M_{\text{NS}} \) should return back toward the BH on timescales equal to or longer than a second. This could in principle produce lower-energy bursts at later times. But as the long-term evolution of the accretion disk is not followed this possibility needs further investigation. Rosswog (2007) returned to this question, considering the accretion activity of material launched into eccentric but gravitationally bound orbits during a compact binary merger coalescence (the main elements of the merger of two compact objects, NS–NS and NS–BH, will be analyzed in Chapter 8). The merger, after a few tens of milliseconds, will leave a remnant that contains the central object, a disk and some material flung into eccentric orbits gravitationally bound to the remnant, what Rosswog calls `fall-back material’. Figure 7.12 (from Rosswog 2007) gives an example of eccentric fall-back trajectories resulting from a double NS merger. The observed X-ray activity might be due to the interaction between the disk and the fall-back material with high eccentricity. The luminosity due to this fall-back material can last hours after the tidal disruption of the NS. Gravitational fragmentation is common in the outer regions of the accretion disks of AGNs (Shlosman, Begelman, & Frank 1990). Fragmentation has also been considered by Perna et al. (2006) and by Piro and Pfahl (2007). If the fall-back material becomes subject to self-gravitational instabilities it may undergo gravitational fragmentation into one or several bound objects. If the fragment is able to impact with nearly free-fall velocity on the disk or if the fragment falls back towards the center and is tidally disrupted, it will be able to trigger violent X-ray flaring activity. Thus, Rosswog considered in his simple model that the late-time X-ray activity might be caused by the fall-back of a small amount (about one-tenth of the disk mass) of highly eccentric material. The fall-back timescales are set by the distribution of eccentricities of this material, considered to be a natural outcome of the merger (Rosswog 2007).

**NS–NS merger creating a millisecond pulsar**

Zhang and Mészáros (2001) considered a continuously injecting central engine with the possibility that this central engine is a highly magnetized millisecond pulsar, a magnetar. Dai and Lu (1998a,b) had previously considered this idea of continuous injection from a millisecond pulsar to interpret pre-Swift afterglow light-curves (GRB 970228). Following the initial fireball (the GRB) produced during the birth of the millisecond pulsar, the authors studied the evolution of a fireball generated post-burst with energy injection from the pulsar, through magnetic dipole radiation. In 2006 Dai et al. revived this model to explain the extended X-ray flares following the short GRBs. The production of a rapidly rotating massive NS (the millisecond pulsar) may be possible if the EOS of the nuclear matter is stiff. The differential rotation leads to a winding up of interior poloidal magnetic fields and the resulting toroidal fields are strong enough to float up and break through the stellar surface. Explosive events driven by magnetic reconnection then occur, which can lead to multiple X-ray flares.
Figure 7.12. (a) Contours of the mass–density \(10^{17} \text{--} 10^{14} \text{ g cm}^{-3}\) 22.3 ms after the merger of two neutron stars with 1.1 and 1.6 \(M_{\odot}\). About 0.03 \(M_{\odot}\) are launched into eccentric orbits and will fall back towards the center at late times (from Rosswog 2007). (b) A sample of fall-back trajectories resulting from the double NS merger with masses 1.1 and 1.6 \(M_{\odot}\) (Rosswog 2007).

minutes after the original gamma-ray burst. This scenario involving millisecond pulsars might explain the X-ray flares observed \(~100\) s after the two short bursts: GRB 050709 and GRB 050724 (Fox et al. 2005, Barthelmy et al. 2005). Dai et al. (2006) also suggested that some long GRBs might originate from moderately magnetized millisecond pulsars with hyper-accreting accretion disks after the collapse of massive stars. The X-ray flares would be the result of the strong interior differential rotation of these pulsars. The difference between the two classes of bursts (long vs short) would be due to differences in the accretion disks: transient torii for short
bursts (Mochkovitch et al. 1993) and fall-back accretion disk for long bursts (McFadyen & Woosley 1999).

**Multistage central engine**

These models are based on the general idea of a short-lived central engine, which collapses into a black hole after some time. We have already presented the model proposed by Staff, Ouyed, and Bagchi (2007) involving a quark star and the delayed formation of a BH. This model can account not only for the energies and the timescales of GRBs but also for the behavior of the early X-ray afterglow. A similar idea was proposed by Gao and Fan (2006) based on a transient magnetar. Other models have been proposed which are referenced and commented on in Zhang (2007a).

In their systematic investigation of the morphological and timing properties of 69 X-ray flares, Chincarini et al. (2007) provide many elements in favor of a common origin of the prompt emission and of the X-ray flares. They point out ten X-ray flares which cannot be explained by refreshed shocks and must be the result of the prolonged activity of the central engine. After three years of Swift operations it was generally considered that the long-lasting activity of the central engine offers the most promising way to explain the new properties of the early X-ray light-curves. The detailed mechanisms involved in this extended activity remain however to be understood.

### 7.5 CHROMATIC X-RAY BREAKS

The lack of early X-ray observations before Swift allowed arbitrary assumptions to be made about the connection between the prompt emission and the afterglow. Usually, the simplest connection between these two phases was assumed, with a possible step at the end of the prompt emission. The existence of the shallow phase was not suspected, and the only known break was the jet break appearing several hours to days after the trigger. Now, there is another break at the end of the shallow phase, at a time $t_b$ around $10^3$–$5 \times 10^4$ s (20 min to 10 h after the GRB). An interesting observation is that this X-ray break is often not seen in the optical (see Figure 7.13, from Panaitescu 2006a), and is thus a ‘chromatic’ break. This chromaticity is, however, limited since there is no spectral evolution of the X-ray emission at the transition between the shallow phase and the pre-jet break power-law phase. In this section we discuss the interpretation of these chromatic breaks revealed by Swift, which occur earlier than the jet breaks discovered by BeppoSAX. The interpretation of the jet breaks in the Swift era will be discussed in Section 7.6.

We have seen that energy injection is the favorite mechanism for explaining the shallow decay phase (see Section 7.2.1). At the end of energy injection an achromatic break is expected, in contrast with the observations as emphasized by Panaitescu
Figure 7.13. Light-curves of six Swift GRB afterglows showing a chromatic X-ray break, not seen in the optical. Optical data are shown with open symbols and are fitted by power laws (dotted lines). X-ray data are shown with filled symbols and are fitted by broken power laws (solid lines). The references to the different measurements are given in Panaitescu et al. (2006a).
(2006). It therefore seems necessary to introduce additional ingredients to explain the chromaticity of the break at the end of the shallow phase. This issue was studied in detail by Panaitescu et al. (2006a) and we follow their analysis below.

The chromaticity of the X-ray breaks indicates that they do not arise solely from a mechanism related to the outflow dynamics (i.e. energy injection) or to the angular distribution of the blastwave kinetic energy (structured outflow or jets). Moreover, the lack of spectral evolution in the X-ray range during the X-ray breaks shows that these breaks do not arise from the passage of a spectral break (most probably the cooling frequency) in the X-ray range. Finally, the decoupling of the X-ray and optical decays implies that the cooling frequency, \( \nu_c \), lies somewhere between the X-ray and the optical domains. Panaitescu et al. (2006a) showed that in this context, the break at the end of the shallow phase must be attributed to the evolution of the cooling frequency due to the end of energy injection. However, this interpretation faces a problem because it implies a post-break decrease of the cooling frequency which is rather fast, \( \nu_c \propto t^{1.3 \pm 0.5} \), much faster than that expected in the standard forward-shock model. In this model under standard assumptions, the steepest decrease of the cooling frequency is \( \nu_c \propto t^{-0.5} \), obtained for a homogeneous medium and synchrotron-dominated electron cooling. The fast evolution of the cooling frequency requires giving up at least one of the standard assumptions. Panaitescu et al. (2006a) proposed considering a variation with time of the two microphysical parameters \( \varepsilon_B \) and \( \varepsilon_e \).

In order to make this evolution more natural, they suggested attributing the evolution of the microphysical parameters to a power-law dependence on the Lorentz factor of the ejecta, \( \Gamma \), combined with a fast decrease of \( \Gamma \) at the end of the shallow phase. Since the shallow phase is believed to be due to energy injection, a change in the dynamics \( \Gamma(t) \) of the outflow at the end of the energy injection seems rather natural. Panaitescu also showed that the evolution of the microphysical parameters \( \varepsilon_e \) and \( \varepsilon_B \) is required, whether the origin of the X-ray afterglow is synchrotron or inverse Compton emission. It has to be noted that the interpretation of pre-Swift afterglows with the standard forward-shock model did not require such an evolution of the microphysical parameters. Interestingly, Panaitescu (2006) noted that, for a steady evolution of the microphysical parameters with \( \Gamma \), and an X-ray light-curve break arising from cessation of energy injection into the forward shock, the six afterglows that he studies require a circum-burst medium with an \( r^{-2} \) radial stratification (i.e., that of a stellar wind). He also noted that the decreasing fraction of shock energy imparted to the electrons helps understanding the apparent high efficiency in the prompt GRB. Of course, as Panaitescu et al. (2006a) recognized, there is no obvious reason for which the microphysical parameters should follow the expected dependence on \( \Gamma \) (\( \varepsilon_e \propto \Gamma^{-i} \), \( \varepsilon_B \propto \Gamma^{-b} \), with \( i/3 + b/4 = 1 \)). In the absence of a physical reason, the authors suggested another possibility, namely that the optical and X-ray afterglow emissions arise from different outflows or different regions of the GRB outflow.

This discussion shows how it is difficult to explain these chromatic breaks with the standard model. In the model proposed by Genet, Daigne, and Mochkovitch (2007), the calculated optical emission from the reverse shock is compared with the
X-ray light-curve (see their Figure 4). Chromatic X-ray breaks are often visible during the first hours, without a corresponding break in the optical band. Uhm and Beloborodov (2007), with a quite similar model (see Section 7.2.6) are also able to produce an X-ray break after the shallow phase and they find that an optical shoulder (not a break) usually accompanies the X-ray plateau. Hence, the reverse-shock model predicts chromatic jet breaks at $\sim 10^5$ s, which appear when the reverse shock reaches the end of the ejecta or faces a steep decline in the ejecta density. Of course, as already mentioned, these two models need much more investigations before revolutionizing the afterglow landscape.

Here again, the chromatic X-ray breaks are very new results, which need more investigation to be fully understood. Nevertheless, this new vision of the X-ray light-curves, with several breaks, some of them chromatic, raises the question of their origin and questions the interpretation of the jet breaks, which should be achromatic. As we will see in the next paragraph, very often this is not the case, leaving open the question of the origin of the jet breaks.

### 7.6 THE JET BREAKS AFTER SWIFT

Different kinds of temporal breaks can be identified in broad-band afterglow light-curves.

#### Jet breaks

Jet breaks have been discussed in detail in Chapters 3 and 6. Being due to a pure hydrodynamical/geometrical effect, they must occur at all wavelengths.

#### Injection breaks

Injection breaks are expected during the early phase of the afterglow, when the total energy in the blastwave is still increasing with time. If the shallow phase is explained by energy injection, the break at the end of this phase can be attributed to the sudden cessation of energy injection. Since this process is also hydrodynamical, this break should also be achromatic and the spectral index across the break should remain the same. We have seen in Section 7.5 that this is not in general the case.

#### Spectral breaks

These temporal breaks occur when a spectral break crosses the observed energy band. Spectral breaks can be due to the characteristic synchrotron frequency $\nu_m$ or to the cooling frequency $\nu_c$. The accelerated electrons may also have an intrinsic break in their energy spectrum (Li & Chevalier 2001). The crossing of the corresponding photon spectral break across the band gives rise to a distinct temporal break. This
type of break must be chromatic moving from high-energy bands (X-rays) to lower-energy bands (optical and radio) (this may not always be the case for wind afterglow models where \( \nu_c \) increases with time (Chevalier & Li 1999, 2000)). For such breaks the spectral indices are different before and after the break.

**Transrelativistic breaks**

At late time, when the fireball changes from a highly relativistic phase to the non-relativistic (Newtonian) phase, there is also a temporal break in the light-curve. Such a break may have been observed in late radio afterglows (see Frail et al. 2003), but without evidence of such a break in optical and X-ray light-curves (Zhang 2007a). For instance, Grupe et al. (2007) reported a long (125 days) X-ray follow-up observation of GRB 060729 that does not show any break in the X-ray light-curve. This steady decay could suggest that the transrelativistic break happens at later times.

These are the most typical breaks, which can be expected in the afterglow light-curves (but the light-curves may be quite complicated, also showing flares and dips).

### 7.6.1 The jet break

The first type of break, which has been discussed in detail, is the jet break introduced very early in the analysis of GRB afterglow light-curves (see Chapters 3 and 6). The beaming of the ejecta was considered as a natural way to avoid huge energy budgets. The jet break time is related to the opening angle, and expected to take place when \( \Gamma^{-1} = \theta_j \), where \( \Gamma \) is the bulk Lorentz factor of the fireball. As explained in Chapter 3, jet breaks have been extensively searched for through the systematic analysis of optical and near-infrared afterglow light-curves. Zeh, Klose, and Kann (2006) have presented one of the most complete analyses of pre-Swift GRBs. They discussed the parameters describing a sample of 16 well-defined afterglow light-curves and the physical quantities related to them. They found that the break times peak at \( \Delta t_b / (1 + z) = 0.3 \pm 0.2 \text{ days} \), without X-ray afterglow measurements to determine whether the break is achromatic. The post-break decay slope is \( \alpha_2 = 2.1 \) but with a broad distribution, and no evidence for a universality of \( \alpha_2 \), as expected if \( \alpha_2 = \beta \) (see also Panaitescu & Kumar 2001, Preece et al. 2002, Yost et al. 2003, for similar conclusions). The distribution of the jet half-opening angle \( \theta_{\text{jet}} \) peaks between 2° and 5°, with a lower cut-off around 2°.

In the pre-Swift era jet breaks were also widely used to establish several empirical relations. The first one was the Frail relation, which allowed defining a beaming-corrected \( \gamma \)-ray energy based on the time of the break, \( t_b \) and on the GRB redshift, and to show that this energy was nearly constant (see Section 3.4). Using this fact, Ghirlanda, Ghisellini, and Lazzati (2004), followed by other authors (e.g. Liang & Zhang 2005, Willingale et al. 2007) found remarkable correlations between quantities related to the prompt phase and the afterglow. The Ghirlanda relation links \( E_p \), the peak energy of the GRB spectrum, and \( E_\gamma \), the beaming-corrected \( \gamma \)-ray energy derived from the break in the optical light-curve (see Sections 3.10 and 9.1).
Independent of the origin of the optical break, Liang & Zhang (2005) found a tight relation connecting $E_p$, $E_{iso}$ (the isotropic equivalent energy), and $t_j$. The relation of Willingale et al. (2007) is different. After the Swift results and the discovery of the shallow phase followed by a break at the end of this phase (see Section 7.2), these authors studied the times of these breaks, already called $t_h$ (this is an X-ray break and not an optical one). The authors noticed a very interesting point: it appears that $t_h$ obtained from X-ray data has properties related to $t_j$ derived from optical data. Even if Willingale et al. (2007) think that it is doubtful that either $t_h$ or $t_j$ are actually jet breaks, it is likely that the end of the shallow phase does depend on the total energy in the outflow, on the collimation angle or on the density of the circum-burst medium, so that the $E_{iso}-E_{peak}-t_h$ relation reported in Willingale et al. (2007) could be related to the $E_{iso}-E_{peak}-t_j$ relation discussed in Liang and Zhang (2005).

Swift, with long and early simultaneous coverage in the X-ray and optical bands, in conjunction with sensitive ground observations, was eagerly expected to identify many achromatic breaks, which would have confirmed the jet scenario. In fact, after the detection of nearly 200 bursts, only a few clear confirmations of achromatic jet breaks can be reported. Figure 7.14 (from Willingale et al. 2007) gives examples of X-ray light-curves in the Swift era. The most common situation corresponds to the two GRBs of the top panels, where the prompt and afterglow components are clearly identified. The middle panels provide examples of the prompt component dominating at late time (left), and a simple component requiring no afterglow (right). The bottom panels give two examples showing a late temporal break at about $10^5$ s.

Even when achromatic breaks are observed, few present the salient characteristics expected for jet breaks. GRB 050801 (Rykoff et al. 2006) and GRB 060729 (Grupe et al. 2007), for instance, have early achromatic breaks covering both the X-ray and optical bands. These breaks, however, correspond to the transition from the shallow to the normal phase, and they are more likely injection breaks.

For GRB 050801 the early afterglow is flat in both optical and X-ray bands and so far unique (see Figure 1 of Rykoff et al. 2006). From their analysis of this burst Rykoff et al. (2006) concluded that the achromatic break visible at $\sim 250$ s cannot be explained by a jet break, and they considered that the early afterglow of this burst is inconsistent with the standard fireball model unless continuous energy injection is involved.

For GRB 060729 the break appears at $\sim 5 \times 10^4$ s. It is visible in $\gamma$-rays and optical light-curves (Figure 8 of Grupe et al. 2007), and its signature in the XRT light-curve is characteristic of the end of the shallow phase. Beyond this break, and up to 100 days after the burst no jet break is detected (this is the longest follow-up and detection of a GRB by the XRT).

For GRB 050525A Panaitescu (2007a) reported the presence of simultaneous breaks in the optical and X-ray light-curves. But the optical emission decays too slowly after the break (index 1.64) to explain the break with a jet, which predicts $\alpha_2 = p$. Moreover, the time of the break has been questioned. Blustin et al. (2006) identified a break at 0.15 days consistent with an achromatic break in the optical and X-ray bands (Figure 7.15), but Della Valle et al. (2006) found a break in the R band at a later time: 0.3 days. Sato et al. (2007) questioned the identification of an achromatic
Figure 7.14. Examples of X-ray light-curves of Swift GRB afterglows fitted with a prompt emission component and an afterglow component. X-ray flares, which are shown, are excluded from the fits. The most common situation corresponds to the two GRBs of the top panels, where the prompt and afterglow components are clearly identified. The middle panels provide examples of the prompt component dominating at late time (left), and a simple component requiring no afterglow (right). This type of light-curve could be due to a ‘naked GRB’ occurring in a very low-density environment leading to the absence of the forward shock and afterglow emission. The bottom panels give two examples showing a late temporal break at about $10^5$ s (Willingale et al. 2007).
break for this burst due to the discrepancy of spectral and temporal relations with the theoretical predictions. The issue remains therefore to be settled for this burst (see Covino et al. 2006).

For GRB 060124 (Curran et al. 2006, 2007), GRB 060526 (Dai et al. 2007; Figure 7.16), and GRB 060614 (Mangano et al. 2007), the observed breaks appear achromatic and the presence of a jet break might be invoked. However, GRB 060124 and GRB 060526 have also been analyzed by Covino et al. (2006). For GRB 060124
these authors noted that even if the achromatic break is quite convincing (see Figure 1 of Curran et al. 2006), the expected post-break temporal decay (equal to $p$ in the X-ray and optical bands), is not observed. The analysis shows shallower and unequal decays in these two wavelength regimes. For GRB 060526, Covino et al. (2006) found that an achromatic break is allowed but not strictly required.

As reported by Panaitescu (2007a), most of the Swift X-ray afterglows well-followed after 1 day do not exhibit a break at 1 day, in contrast with the well-followed pre-Swift optical afterglows. For a sample of twenty-five X-ray afterglows extending beyond 5 days, only eight show a second steepening break at 0.1–3 days, which might be interpreted as a jet break. For two other events, GRB 050318 and GRB 060124 (but see the previous discussion on this GRB), only one steepening break is observed at 0.2–2 days, with the post-break being sufficiently steep to warrant a jet break interpretation. Optical measurements before and after the X-ray breaks exist for four of these ten GRBs: GRB 050525A, GRB 051221A, GRB 060124 and GRB 060526 (we have already discussed three of these bursts above). Figure 7.17 (from Panaitescu 2007a) gives the optical and X-ray light-curves of three of them. Potential jet breaks in the optical light-curves at 1 day are reported, but for GRB 050525A and GRB 051221A the optical emission decays too slowly after the break (with indices of 1.64 and 1.20, respectively) to be consistent with a jet break interpretation. The conclusion of Panaitescu (2007a) is that for the X-ray emission of three dozen Swift afterglows followed for several days to weeks after the burst, three-quarters of them do not show a steepening at 1 day, i.e. at the time where most of the pre-Swift optical breaks were observed. Only four of them provide some evidence for an achromatic steepening, with the reservations that we have indicated. So, Panaitescu (2007a,b) concluded that there is a lack of evidence for achromatic breaks and recommends a more efficient optical monitoring of future Swift afterglows to test the prediction of the widely used jet model.

Zhang (2007a) arrived at the same conclusion: the present Swift data do not seem to support the existence of jet breaks, casting doubt on the interpretation of the pre-
Swift breaks. As a last example illustrating this remark we can mention the optical light-curve of GRB 060206, which exhibits a late-time steepening of the light-curve, suggestive of a jet break. Unfortunately, the X-ray data show a remarkable single power-law decay without any evidence of a break at the time of the optical break (Monfardini et al. 2006, Burrows & Racusin 2007). Along the same lines, various authors have tried to use the Ghirlanda relation to ‘predict’ the time of the X-ray breaks from the measure of $E_p$ and $E_{iso}$ in the prompt emission. Sato et al. (2007) have searched for expected jet breaks of three Swift bursts (GRB 050401, GRB 050525A, and XRF 050416A) in the X-ray band without results. For this last XRF Mangano et al. (2007b) reported the lack of a jet break in the X-ray light-curve, which is measured by the XRT for $\sim 6 \times 10^6$ s (Figure 7.18). Willingale et al. (2007) extended this search and gave eight examples of X-ray light-curves in which a ‘predicted’ jet break is absent (Figure 7.19). Covino et al. (2006) summarized the search for achromatic breaks in Swift afterglows using high-quality multi-wavelength data and concluded that no convincing case is identified. So, the very origin of the observed breaks is still a mystery and it is not firmly established that the pre-Swift ‘jet breaks’ are indeed jet breaks, because their achromaticity could not be robustly established. Most of the pre-Swift jet breaks were observed in one band only, usually in optical and sometimes in X-rays and radio, and even the best case, GRB 990510 (Harrison et al. 1999; Figure 3.9) is not so sure (see the comments of Zhang (2007a)).

Having raised the question of the nature of the observed breaks, Zhang (2007a) notes that X-ray breaks seem to systematically lead optical breaks which in turn systematically lead radio breaks. This observation drives him to propose a scenario in which these temporal breaks are spectral breaks that roll down from high energies to low energies. Initially the break is in the gamma-ray band and it defines $E_p$ in the prompt emission spectrum. Later on ($10^3–10^4$ s) this break moves to the X-ray band, giving rise to the early X-ray breaks observed in some bursts. Then, around 1 day, this
break rolls down to the optical band and it can account for the pre-Swift optical breaks that are interpreted as jet breaks. It finally moves to the radio band at \(\sim10\) days. Zhang (2007a) used the example of GRB 030329 (see Section 4.3), with an early optical break and a later radio break (Berger et al. 2003) in support of this model. With this scenario the link between \(E_p\) and the optical and X-ray break times (as evidenced by the Liang & Zhang and Willingale relations) is easier to understand. Even though this idea is attractive there are difficulties, as recognized by Zhang (2007a). The most severe is that one expects changes of the spectral index across the breaks, which are not observed for most bursts. Another problem is that in such a scenario the prompt emission and the afterglow must come from the same emission component. But Swift seems to have demonstrated through the studies of early afterglows that the prompt emission and the X-ray flares are due to internal shocks while the afterglows are not.

This discussion clearly indicates that the question of breaks in afterglow light-curves remains an important and puzzling open issue. Ironically, while the new X-ray observations of Swift provide an extensive coverage of the X-ray afterglow, it is now

**Figure 7.18.** The beautiful X-ray light-curve of XRF 050416A. The light-curve has the general characteristics reported for GRBs and no jet break is seen beyond the end of the shallow phase in spite of a very long monitoring up to 74 days after the burst. The phenomenology of this XRF is a supplementary element to support the idea that both GRBs and XRFs arise from the same phenomenon. Jet breaks are as rare for XRFs as for GRBs (from Mangano et al. 2007b).
Figure 7.19. X-ray light-curves in which a jet break is not seen at the epoch predicted by the Ghirlanda relation. $E_{\text{peak}}$ is in kiloelectronvolts, $E_{\text{iso}}$ in $10^{52}$ erg, and $t_j$ is in days. The shaded regions indicate the time range where a jet break should appear if GRBs were to have a constant reservoir of beaming corrected energy (from Willingale et al. 2007). In GRB 050820A, the prompt emission peaks between 200 and 300 sec, explaining the huge flare seen at that time.

in the optical band that the observations are often too scarce. These observations are crucially needed to prove or disprove the interpretation of optical breaks as jet breaks. A related point is the importance of using the large database produced by Swift to determine global statistical properties which are now essential to making progress in our understanding of GRBs (see, for instance, the studies of Willingale et al. 2007, Chincarini et al. 2007, Zhang et al. 2007, Zhang, Liang, & Zhang 2007, Panaitescu 2007a,b). These detailed observations may lead to dramatic revisions of the simple internal/external-shock model, as it was developed to explain BeppoSAX observations. We have cited some of them in this chapter. The central engine may
remain active for hours or days, injecting energy in the ejecta with all kinds of Lorentz factors. The role of the reverse shock could be much more important than was previously thought (Genet, Daigne, & Mochkovitch 2007, Uhm & Beloborodov 2007). The microphysical parameters may vary, following the evolution of the ejecta (Panaitescu 2006). The ‘jet breaks’ may not be simply related to the beaming of the ejecta. All these issues will be explored in the coming years, allowing the next chapters in the history of GRBs to be written, with the new details and reinterpretations, which will continue to make these events so fascinating.

7.7 CONCLUSION

In this chapter we have analyzed how the major results of Swift on the early X-ray afterglow can be explained in the framework of the fireball model. We have also discussed the crucial question of jet breaks linked to the dearth of achromatic breaks in a significant sample of GRBs.

We have seen how the two major discoveries of Swift concerning the early X-ray afterglow—the steep decay and the shallow phases—can be explained, with significant consequences for the GRB models. The steep decay phase can be attributed to the so-called curvature effect, but time-resolved spectral analyses of the tails of 44 GRBs, the majority of which display hard-to-soft spectral evolution, are not easily explained by the curvature effect. So this explanation cannot be considered as definitive. The shallow phase is a new major concern. As we have indicated many ways have been explored to explain it, among them: different jet models, prior activity, and time-dependence of the microphysical parameters. But the more popular scenario is based on late energy injection requiring long-lived activity of the central source, and another possibility is a delayed energy transfer to the forward shock. The discovery of this shallow phase has necessitated revisiting the efficiency of internal shocks, especially for the leading scenario of late energy injection. Despite large uncertainties, the shallow phase may eventually endanger the internal-shock model. In fact the difficulties that have followed the discovery of the shallow phase have led to some proposals of dramatic revisions of the classical fireball model with internal and external shocks. We have briefly described new ideas which might resolve these difficulties. Among them it is proposed to attribute the early afterglow and possibly the whole afterglow to a long-lived reverse shock in the ejecta. This requires specific conditions: a radiatively inefficient forward shock, a broad range of Lorentz factors in the ejecta allowing the reverse shock to remain active for long times, and the acceleration of only a fraction of the electrons at the shock. Today the explanation of the shallow phase is largely an open issue, even though the late energy injection is the most accepted model.

A second fundamental discovery of Swift was the presence of X-ray flares in many GRBs (about half of them). Again, the standard fireball model did not anticipate such flares. Many authors have concluded that a large fraction (if not all) of X-ray flares are due to long-lasting activity of the GRB engine rather than to external shock activity. Models involving internal shocks have several advantages
because the X-ray flares and the underlying X-ray afterglow observed at the same observer time are in fact originating from different physical sites at different times. Hence, as for the shallow phase, the X-ray flares seem to require late central engine activity in their more realistic models.

The next step was to explore how this long and/or late central engine activity could be achieved. We have examined different suggestions: fragmentation or gravitational instabilities in the massive star envelope or in the accretion disk, a magnetic barrier around the accreter, NS/NS mergers with the formation of a millisecond pulsar. Today the long-lasting activity of the central engine offers the most promising way of explaining the new properties of the early X-ray afterglows discovered by Swift. But, even though several possibilities exist, a lot has still to be done to prove that current models can account for this extended activity.

Swift has also contributed decisively to the question of jet breaks. The optical breaks observed in the pre-Swift era were considered as well-established jet breaks, but their achromaticity was not really demonstrated in an extended range of wavelengths. Swift enabled the first long-duration X-ray observations starting very early after the burst, and the detailed comparison of the X-ray and optical light-curves of the afterglow. This comparison has shown that the achromaticity of breaks was more the exception than the rule. In most cases no steepening of the X-ray light-curve is observed at 1 day, at the time at which most of the pre-Swift optical breaks were reported. Even when achromatic breaks in the optical and X-ray ranges are observed, their interpretation as jet breaks is often questionable. Hence, the very origin of these breaks is still a mystery and it cannot be established for sure that pre-Swift optical breaks were jet breaks, with all the consequences that this interpretation has led to in the BeppoSAX era.

Finally, Swift revealed definite X-ray breaks at the end of the shallow phase but these breaks occur earlier than the optical breaks discovered during the BeppoSAX era and they are not associated with similar steepening of the optical light-curves. The chromaticity of these breaks indicates that they do not arise solely from a mechanism related to the outflow dynamics. More investigations are needed to understand the origin of these X-ray breaks.

In conclusion, even though Swift is returning remarkable results, we can say that today the explanation of the GRBs seems to be more complicated than it appeared to be after the first BeppoSAX discoveries. The mechanisms that explain the origin and the development of GRBs are still obscure. In the next chapter we will see that, in spite of beautiful ideas and deep searches, our understanding of the progenitors which can generate such explosions also faces various difficulties. This understanding is complicated by the fact that the GRB emission is released far from the central engine, this being particularly true for the afterglows.

7.8 REFERENCES


Derishev, E. V., Kocharovsky, V. V., & Kocharovsky, V. V. (1999b) Lightcurves of cosmological gamma-ray bursts, *Astronomy and Astrophysics* 345, L51–L54.


8

Progenitors

8.1 INTRODUCTION

We have seen that, thanks to BeppoSAX, the cosmological nature of GRBs is now well established. The observed fluences and the determination of redshifts have made it possible to measure the isotropic equivalent energies emitted by GRBs in.gamma-rays. These energies reach $10^{50}$ to $10^{54}$ erg. As most GRBs are probably collimated with typical opening angles between 1° and 20°, the true energy is clustered around $10^{50}$ to $10^{51}$ ergs which must be liberated in a matter of few seconds. These values are smaller than the binding energy of a stellar compact object. The central engine must be able to generate such energies and accelerate typically $10^{-5} M_{\odot}$ to relativistic velocities. We have seen that, in spite of some recent difficulties, the fireball model with its internal and external shocks is today the most successful at explaining a lot of the properties of GRBs and their afterglows. The GRB variability timescales suggest that the energy is quickly deposited by a central engine within a small volume, a few hundred kilometers, which is optically thick to gamma-ray radiation at early times. This opaque fireball then expands adiabatically and relativistically. To reach ultra-relativistic regimes this fireball has to have low baryon loading or to be dominated by magnetic energy (Poynting flux). This is the condition for reaching high Lorentz factors for the expanding fireball, typically one hundred to a few hundred. This adiabatic expansion will stop when the gamma-ray radiation is able to escape, and this happens between $10^{11}$ and $10^{12}$ cm from the explosion site (see Figure 5.1). The GRB is certainly due to internal shocks: the interaction within the ejecta of relativistic shells produced by the central engine.

As the variability timescale can be as short as 1 ms and the overall duration of the GRB can be tens of seconds, the internal-shock model requires a central engine which has to be compact, with prolonged activity corresponding more or less to the GRB duration (see Chapter 5). This rules out any explosive model in which the energy is released instantaneously (Piran 2005). The main characteristics of the GRBs and their
afterglows are reproduced by the fireball model, which includes the ultra-relativistic expansion of the fireball and its interaction with the interstellar medium, and the emission of radiation by synchrotron and/or synchrotron plus inverse Compton processes. Now a major question is still open: What makes the GRB and what makes the ultra-relativistic fireball?

What are the progenitors, the astrophysical objects capable of producing these huge explosions? As we have said, the total energy released is certainly below the rest-mass energy of stellar mass objects. The short variability time implies energy deposition in a very small volume, so it appears natural to invoke compact objects and probably black holes. Moreover the GRB occurrence rate and the lack of repetition suggest that these events are rare. Depending on the collimation of the fireball, typical GRBs would have to take place once per $10^4$–$10^6$ years per galaxy, about 1% to 0.01% of the rate of supernovae (Piran 2005). Since no repetition has been established it means that during these events the progenitor is catastrophically destroyed.

As the energy is not released instantaneously, it has been suggested that GRBs are associated with astrophysical objects involving accretion of a massive ($\sim 0.1 \, M_\odot$) disk onto a compact object, most likely a newborn black hole. The idea of a compact object is connected with the short variability time. The energetics requires the presence of a massive disk which has to be formed simultaneously with the formation of the compact object.

Such a system can result from the explosion of a massive star, as in the collapsar model, or following the coalescence of binary compact stellar remnants, involving neutron stars (NSs) and/or black holes (BHs). In both cases a spinning black hole is formed and the debris, either from the stellar core of the collapsar or from a tidally disrupted neutron star, form a temporary accretion disk or torus. This material will ultimately fall into the black hole, releasing a fraction of its gravitational energy. The GRB can thus be powered by the gravitational energy of the in-falling matter. The spin energy of the black hole can be another source of energy through the Blandford–Znajek process (Blandford & Znajek 1977). An alternative to accretion was proposed by Usov in 1992. It involves pulsar-like activity of a rapidly rotating compact object with a very high magnetic field: a newborn ‘magnetar’.

In this chapter we will discuss extensively the various kinds of progenitors that have been proposed to explain all types of GRBs. We will start with collapsar models thought to explain long GRBs in Sections 8.2 to 8.4, and quickly discuss pulsar models in Section 8.5. Section 8.6 will review the observational evidence linking long GRBs with collapsars. We will then turn to short GRBs ($T < 2 \, \text{s}$; see Chapter 2) and discuss at length the possible ways of making binary systems of compact objects that will end up coalescing and making a short GRB.

8.2 THE COLLAPSAR MODEL

Collapsars are rotating massive stars either isolated or in a binary system whose iron core eventually collapses directly to form a black hole (Woosley 1993). In the
seconds/minutes following the collapse, the black hole accretes the residual matter of
the core and emits a powerful relativistic jet that will be detected as a GRB if directed
towards the Earth. To avoid a significant baryon loading the progenitor star should
have lost most of its extended gas envelope of hydrogen by the time of the collapse.
The progenitors of collapsars, likely Wolf–Rayet stars, are thus closely related to
the progenitors of hydrogen-deficient supernovae, i.e. type Ib/Ic supernovae
(MacFadyen & Woosley 1999). A complete discussion of the possible stellar progenitors
leading to a black hole with an accretion disk can be found in Fryer et al. (1999; see their various scenarios in Figure 8.1). The collapsar model requires three
essential ingredients, of which we have already considered two: a massive core and the
removal of the hydrogen envelope (we will see why below). The last condition is a
high angular momentum in the core, which is needed to support a transient torus
around the black hole resulting from the collapse of the massive core. The spinning
black hole is important to achieve an efficient energy conversion and to create a
natural rotation axis free of matter, along which a jet and an expanding blastwave
will be able to escape.

To give the main characteristics of the collapsar model we will use the detailed
analysis of MacFadyen and Woosley (1999), this model having been introduced for
the first time by Woosley (1993). Among the numerous publications which have
followed we have selected those of Fryer, Woosley and Hartman (1999), MacFadyen,
Woosley, and Heger (2001), Zhang, Woosley, and MacFadyen (2003), Zhang,

In its initial version of the collapsar model Woosley (1993) considered stars
heavier than about 35–40 $M_\odot$ which are assumed to lose their hydrogen envelope
before their death (Chiosi & Maeder 1986). These large stars have big iron cores,
which make them more difficult to explode (Wilson et al. 1986). Such stars with
masses around 25–35 $M_\odot$ on the main sequence and with a helium core of 9–14 $M_\odot$
might be good candidates to become what Woosley called ‘failed supernovae’: they
do not explode, resulting in the prompt formation of a black hole with a massive
accretion disk, and opening the possibility of producing a GRB. The study of ‘failed
supernovae’ was initiated by Bodenheimer and Woosley (1983) and explored in a
preliminary way as GRB progenitors by Woosley (1993, 1996) and by Popham,
consequences of the collapsar model in a phenomenon he called ‘a hypernova’.

MacFadyen & Woosley (1999) explored in particular the fate of a 14 $M_\odot$ helium
core from a 35 $M_\odot$ main sequence star. We summarize here the main conclusions
that they have obtained using a two-dimensional hydrodynamics code to study the
evolution of such stars (for a complete analysis, see their paper and references
therein). The key element of the model is the two-step collapse of the iron core of
a rapidly rotating helium star, first to a neutron star, then promptly to a black hole,
because an outgoing shock fails to be launched by the collapsed iron core. In this
model, which will be later called a type I collapsar (MacFadyen, Woosley, & Heger
2001) the black hole forms within a few seconds of the collapse of the iron core, while
the in-falling matter in the equatorial plane is slowed by rotation and piled into
disk. No outward-moving shock is generated prior to disk formation. A very
Figure 8.1. Three scenarios of collapsar formation proposed by Fryer, Woosley and Hartmann (1999), involving either a single star or binary stars. Scenario X: a single rotating Wolf–Rayet star is involved, with winds blowing off its hydrogen mantle. A massive helium core is left, which then collapses into a black hole (BH). This core must have a mass greater than 10 M⊙ to insure the immediate formation of a BH without SN explosion. Scenario XI: this is the dominant collapsar scenario. The hydrogen mantle of a rotating massive star is driven off during a common envelope phase. The helium core then collapses to form a black hole. The final outcome can be a BH/WD binary or a BH/NS binary. Scenario XII: a double He star is formed after a common envelope phase, the two He stars merge and the collapse leads to a GRB. This scenario (Brown 1995) can also lead to a double NS binary and eventually to a double NS merger.
energetic stellar explosion (up to $\sim 10^{52}$ erg) is observed powered by hyper-accretion into the black hole. As a consequence of the in-fall into the black hole of the matter which was initially situated along the rotational axis and of the stagnation of matter in the equatorial disk, a favorable geometry for jet outflow develops. The first important stage for the evolution of this star is the formation of the disk and the partial evacuation of polar regions. This takes several seconds because the jet cannot develop as long as polar accretion continues at a high rate. After several seconds the situation changes, the pole is sufficiently clean for a reversal of the flow to become possible. Energy is dissipated in the disk by neutrino annihilation which can power polar outflows, relativistically expanding bubbles of radiation, and pairs and baryons focused by density and pressure gradients into jets. Since the black hole rotates very rapidly, and because strong magnetic fields are expected to develop in the disk, MHD processes may also contribute to energy dissipation (the rotational energy of the black hole can also be extracted via the Blandford–Znajek mechanism—1977).

So the jet may be energized either by neutrino energy transport or by MHD processes (or both, for instance the jet might be initially dominated by thermal energy and make an early transition to become dominated by magnetic energy—a Poynting flux—or the reverse). MacFadyen & Woosley (1999) found that the total neutrino energy emitted by the disk during the 20 s of the accretion phase might reach $3 \times 10^{33}$ erg and the total energy deposited by neutrino annihilation along the rotational axis would be $1-14 \times 10^{51}$ erg. But these numbers are obtained for the particular model considered by the authors. A typical factor-of-10 uncertainty can be due to the sensitivity of the model to the accretion rate and to the Kerr parameter characterizing the initial angular momentum of the black hole, but also to the uncertainty in the neutrino efficiencies. Larger values are possible especially if the accretion rate is a little higher. Energy deposition is simulated with its consequences: a highly focused relativistic outflow and jet development. The energy deposition blows a low-density bubble, avoiding baryonic contamination. Momentum and energy from the annihilating neutrinos continue to be deposited in this bubble where the pressure gradient tends to exclude gas from the polar region. The energy is naturally directed outward along the axis and the expanding region remains very elongated, ‘focused’. The jet, which is followed through a significant fraction of the star’s mass and radius, maintains a collimation with a half angle of about $10^\circ$.

This study shows that the jet is capable of breaking out of the star in $\sim 10$ s, maintaining collimation and relativistic speeds and clearing a channel through the star for the unhindered escape of what is essentially a pair fireball.

As the jet breaks through the surface mildly relativistic matter is ejected and further accelerated by shock steepening in the density gradient. As this matter runs into the previously ejected wind of the star a relatively weak soft GRB may be created (Woosley, Eastman, & Schmidt 1999). This may also be the origin of hard X-ray precursors seen in some GRBs. The GRB itself appears only after the jet has broken out of the star and continued long enough to clean a low-density channel for the relativistic plasma. Typically, the GRB starts $\sim 10$ s after the black hole forms and it is produced far away from the star, at about $10^{14}-10^{15}$ cm, when the jet has already become optically thin. The authors observe that the accretion rate is not steady...
during the burst, and since the efficiency of energy deposition is sensitive to this accretion rate there is an instability leading to highly variable energy deposition in the polar regions. Disk instabilities can thus give rise to variability on all timescales up to a few tenths of a second. Part of this variability may persist in the time structure of the burst. Finally, the jet can turn on and off on timescales of 50–300 ms and it is expected that the energy of the jet, i.e. its Lorentz factor $\Gamma$, could be highly variable. GRBs produced by internal shocks in the jet are thus favored.

MacFadyen and Woosley (1999) noted that collapsars are unable to produce bursts shorter than $\sim 5 s$. Bursts shorter than 5 s can only be the result of ‘seeing the tip of the iceberg of longer bursts’. The existence of short hard bursts with a mean duration of a few tenths of a second seems to require other progenitors, for instance coalescing compact objects.

Concerning the collapsar model MacFadyen and Woosley (1999) predicted that all GRBs produced by this model will also make supernovae like SN 1998bw. So, even if a small fraction of SN makes GRBs, it seems that most long GRBs make a SN. Of course the optical afterglow from the deceleration of the jet is so bright that it may outshine the fainter supernova. In conclusion it is clear that collapsar progenitors of GRBs should be directly associated with star-forming regions, which are the place where these massive stars are born and where they die. After their original paper, this group has carried out more detailed studies of collapsars (MacFadyen, Woosley, & Heger 2001, Zhang, Woosley, & MacFadyen 2003, Zhang, Woosley, & Heger 2004).

MacFadyen, Woosley, and Heger (2001), continuing the study of GRB produced by the collapsar model, introduced another type of collapsar they called type II. In this case the black hole is formed over a longer period of time in a mild explosion, i.e. a SN that ejects an outgoing shock that does not have enough strength to expel all the helium and heavy elements outside the neutron star (Fryer 1999). He estimates that this behavior might involve stars with masses between $\sim 20$ and $40 M_\odot$. Figure 8.2 (Fryer 1999) indicates three main mass ranges: below $\sim 20 M_\odot$ a neutron star is formed, between 20 and $40 M_\odot$ the ejected material decelerates and falls back onto

**Figure 8.2.** Binding energy (solid line) and explosion energy (circles) as a function of the mass of the progenitor. This binding energy includes all but the inner $3 M_\odot$ core of the star. If the explosion energy is less than the binding energy, the compact remnant will exceed $3 M_\odot$ and collapse to form a BH. The explosion energy drops with increasing progenitor mass, while the binding energy rises. These two effects limit the mass transition between NS formation and (fall-back) BH formation to a narrow range. Vertical lines delimit the three domains of NS formation, fall-back BH formation, and direct BH formation (from Fryer 1999).
the neutron star leading to the delayed formation of a black hole (Woosley & Weaver 1995), above $\sim 40 M_\odot$ a black hole is formed directly with no SN explosion. For type II collapsars the infall of 0.1 to several $M_\odot$ of the stellar mantle would be able to produce a black hole and part of the accreted mass will form a disk if it has sufficient angular momentum. The end result of the explosion will be a black hole with a torus of matter as in a type I collapsar. Type II collapsars with delayed black hole formation must be more frequent than type I collapsars because they involve a larger population of stars. MacFadyen, Woosley, and Heger (2001) showed that in this model the typical accretion rate when most of the matter falls back is much lower than in type I collapsars (one or two orders of magnitude less). This is too low for neutrinos to extract disk energy effectively and form a jet (Popham, Woosley, & Fryer 1999) but such accretion rates are similar to those invoked for magnetohydrodynamic (MHD) models of GRBs. This has motivated MacFadyen, Woosley, and Heger (2001) to consider the possibility that energetic jets will form in a SN that is already in the process of exploding. For these low accretion rates the energy of the jet and the explosion it produces depend upon the efficiency of MHD processes in extracting energy from the disk and the black hole. As explained below, several processes may operate in the vicinity of the rotating black hole to create a jet.

- Very close to the black hole, magnetic fields maintained by currents in the disk and threading the black-hole ergosphere can extract the rotational energy of the black hole, this is the Blandford–Znajek mechanism (1977). This effect appears when there are sufficient charges around the black hole to provide the force-free condition. Wald (1974) argued that when a rotating black hole is situated in an externally supported field in vacuum, an electric field is induced and the lowest energy state has a finite charge on the hole. In this situation there is no radiation of energy. However, in general, the electric field has a non-zero component along magnetic field lines. Blandford and Znajek (1977) (BZ) showed that if the magnetic field and the angular momentum are large enough, the vacuum surrounding the hole is unstable because charged particles will be electrically accelerated and will radiate. This radiation will produce further charged particles in the form of $e^+e^-$ pairs. When charges are produced so freely, the electromagnetic field in the vicinity of the horizon will become approximately force-free (Blandford & Znajek 1977); see for instance the schematic picture of a force-free magnetosphere given by Okamoto (1992). Under these conditions the magnetic field lines exert no force and corotate rigidly with the rotating black hole. The induced current loops which pass along the black hole’s stretched horizon feel the forces due to the magnetic field supported by the surrounding material (the accretion disk). These forces provide the magnetic braking of the black hole rotation (Lee, Wijers, & Brown 2000). The presence of this accretion disk is important for the BZ process because it is the supporting system of the strong magnetic field threading the black hole, which would disperse without the pressure from the field anchored in the accretion disk (see, for instance, Figure 3 of Hirotani et al. (1992) for a schematic view of the field lines connecting the horizon with the surrounding disk). A black hole by itself cannot keep the
magnetic field on it for a long time. So there exists a magnetic coupling between the BH and the disk and, as described by Lee, Wijers, and Brown (2000), the magnetic field of the disk extracts not only angular momentum but also a substantial energy out of the disk as it does with the BH. Concerning the BH we can refer to the situation considered by Lee, Wijers, and Brown (2000) with a half hemisphere rotating with angular velocity \( \Omega \) and a circle on the surface across which a surface current flows down from the pole. Since the external magnetic field must thread the surface outward normally, any surface current feels a force, and a torque results due to Lorentz forces. This torque is the source of the magnetic braking; the rotational energy loss which results is proportional to the torque and to \( \Omega \). In the case of a black hole, these quantities \( \Omega \) and \( B \) are defined on the stretched horizon (Lee, Wijers, & Brown 2000). So a huge energy can be extracted from a rotating black hole through the BZ process. For a maximally rotating black hole, the rotational energy which can be extracted is 

\[
0.29\varepsilon M_{\text{BH}} c^2 \sim 5 \times 10^{43} \varepsilon (M_{\text{BH}}/M_\odot) \text{ erg},
\]

where \( \varepsilon \) is the efficiency in converting the rotational energy into MHD jet energy (Mészáros, Rees, & Wijers 1999). The maximum efficiency is \( \varepsilon = 31\% \) (Lee, Wijers, & Brown 2000), hence a maximum of 9% of the rest mass energy of the black hole can be extracted.

- Farther out, other MHD processes may also extract angular momentum and energy from the disk to form a jet, for instance by magneto-centrifugal forces (Blandford & Payne 1982, Koide, Shibata, & Kudoh 1999, Meier 1999). These processes tend to produce relatively cold jets, i.e. the thermal energy of the jet is not initially large compared with either its rest mass or its kinetic energy. On the other hand other processes such as neutrino energy deposition (Woosley 1993, MacFadyen & Woosley 1999) or magnetic reconnection in the disk (Thompson 1994, Katz 1997) deposit their energy chiefly as heat and make hot jets. Cold MHD jets may be collimated both by magnetic fields and geometry but hot jets (those considered in MacFadyen & Woosley (1999) for type I collapsars) are collimated only by the structure of the medium in which they expand (MacFadyen, Woosley, & Heger 2001). For instance the hot jets shown by MacFadyen and Woosley (1999) are focused by the thick accretion disk.

Coming back to the collapsars, MacFadyen, Woosley, and Heger (2001) compared the observable phenomena resulting from collapsars of types I and II. Considering three cases: with no hydrogen envelope (Wolf–Rayet (WR) stars), with a small hydrogen envelope (blue supergiant), and with a large hydrogen envelope (red supergiant), they give in their Table 3 the type of SN explosion and the type of GRB or X-ray transient which can be expected. Again, we note the importance of the stellar envelope for the collapsar evolution and the burst which can result. Aloy et al. (1999) have recalculated the 14 \( M_\odot \) collapsar model of MacFadyen & Woosley, using a 3D, fully relativistic version of their code and they obtain similar results. The relativistic jet forms as a consequence of an assumed energy deposition (\( 10^{50–51} \text{ erg s}^{-1} \)) within a 30° cone around the rotation axis. The jet flow is strongly beamed (few degrees) spatially inhomogeneous and time-dependent. The jet is able to reach the surface of the stellar progenitor (typically \( R \sim 3 \times 10^{10} \text{ cm} \)) intact. At breakout, the Lorentz
factor of the jet reaches $\Gamma \sim 33$. After breakout the jet accelerates into the circumstellar medium. At $R' = 2.54R$, the end of the simulation, about 2 s after shock breakout, the Lorentz factor has increased to $\Gamma = 44$ in the core of the jet which is now highly collimated ($\sim 1^\circ$). At that time the jet has reached $7.5 \times 10^{10}$ cm ($10^2$ to $10^4$ smaller than the distance at which the fireball becomes optically thin) and the values of $\Gamma$ are still far from those required for the fireball model (Aloy et al. 2000). These simulations also show that within the jet the baryon load is very inhomogeneous, and the authors conclude that the concept of $\eta$ (the energy-to-mass ratio in the fireball; see Chapter 5) as a global parameter is useless. They also note that Lorentz factors of several hundred might be reached before the central engine is switched off.

Two later papers by Zhang, Woosley, and MacFadyen (2003) and Zhang, Woosley, and Heger (2004) are interesting because they go a little farther in the simulations. While the simulation stops when $\Gamma$ reaches 44 in the work of Aloy et al. (2000), in Zhang, Woosley, and MacFadyen (2003) the conditions of jet formation are not considered, and the jet is initiated in a parametric way based on its power, opening angle, initial Lorentz factor and internal energy. The propagation of the jet to the stellar surface at $\sim 8 \times 10^{10}$ cm and its interaction with the stellar mantle are followed. The calculations also examine what happens to the jet immediately after it escapes the star and converts its residual internal energy into additional relativistic motion. Like Aloy et al. (2000), Zhang, Woosley, and MacFadyen (2003) observed that the passage of the jet through the star leads to an additional collimation: starting with an initial half angle of 20$^\circ$ the jet emerges with a half angle of 5$^\circ$. Moreover, instabilities along the beam’s surface lead to mixing with the nearby stationary stellar material and to possible sporadic decelerations of the flow. The mixing produces variations in the mass loading and therefore the Lorentz factor of the jet is variable as is needed in the internal-shock model. The opening angle of the jet varies with time, growing as the star is blown aside. This implies that what the observer will see will be quite sensitive to the viewing angle. The jet that finally emerges from the star has a moderate Lorentz factor, which is modulated by mixing, and a very large internal energy. Zhang, Woosley, and MacFadyen (2003) followed the escape of the jet and observe that the conversion of the remaining internal energy leads to a terminal Lorentz factor along the axis of approximately 150, for the conditions they consider. As an example their Figure 14 (Figure 8.3) gives the Lorentz factor at $t = 10$ s and $t = 35$ s for one of their models. The opening angle of the final jet can be significantly greater than at breakout. A small amount of material emerges at large angles, with a Lorentz factor still large enough to make a weak burst. This suggests a unified model which can explain various types of bursts, from X-ray flashes to classical GRBs, with the same collapsar model, the distinction between the different bursts being due to the viewing angle (Zhang, Woosley, & MacFadyen 2003). With this model GRB outflows have a narrow and highly relativistic jet beam and a wide, mildly relativistic jet wing. This supports a non-uniform jet model, which was proposed for instance by Rossi, Lazzati, and Rees (2002) and Salmonson (2003) and discussed in Section 6.7.4.

To complete this short review of collapsars and their jet production it is interesting to refer to the paper of Zhang, Woosley, and Heger (2004). They presented
new 2D and 3D calculations of relativistic jet propagation and breakout in massive WR stars. These two kinds of calculation lead to an important conclusion which is robust: a relativistic jet can traverse a WR star while keeping sufficient energy and high Lorentz factor to make a GRB. Thus, the collapsar model can be considered as secure. Moreover, as we have seen, the typical relativistic jet core is collimated to 3° to 5° with a Lorentz factor ∼100. The core of the jet is surrounded by a cocoon of less energetic and moderately relativistic ejecta (Γ ∼ 15) that expands and is visible at larger polar angles (∼10°). The word cocoon had already been used by Scheuer (1974) to designate mildly relativistic material that surrounds the relativistic jet of a radio-source. This mildly relativistic cocoon can be the origin of transients accompanying the GRB, X-ray flashes, and other high-energy transients, by its emission at lower Lorentz factors and for angles about three times greater than the GRB. They would be observable from a larger fraction of the sky but, as they are weaker than typical
GRBs, they would be visible at shorter distances (a median redshift of $z < 1$ is suggested). In any case, if the model is valid, XRFs and GRBs should be different expressions of the same basic phenomenon. XRFs are expected to have an energy which is only a few percent of GRBs and to be about 1% as frequent in the observed sample (Zhang, Woosley, and Heger 2004).

The question of X-ray precursors is also interesting in the context of the collapsar model. MacFadyen and Woosley (1999) have considered the possibility of having an X-ray precursor when the jet first breaks out of the star. The possible photon signatures of the jet as it emerges from the massive progenitor have also been considered by Waxman and Mészáros (2003). They introduced what they called an ejected cork (see their Figure 2). This cork, which caps the jet that makes its way through the star, is composed of a termination shock advancing into the star and a reverse shock moving back into the jet. Initially this cork is optically thick, and it experiences successive shocks as the optical depth decreases during its overall expansion. This behavior leads to an X-ray precursor to the usual GRB. Such X-ray precursors were also inferred by Ramirez-Ruiz, MacFadyen, & Lazzati (2002).

Another way to produce $\gamma$-ray and X-ray transients has been proposed by Ramirez-Ruiz, Celotti, and Rees (2002), and by Mészáros and Rees (2001). It is the consequence of the behaviour of the cocoon that surrounds the jet. When the jet makes its way through the stellar core its progression speed is slowed down and most of the energy output during that period is deposited into a cocoon of relativistic plasma surrounding the jet (Mészáros & Rees 2001). This plasma is able to escape along the rotation axis of the progenitor and to accelerate in approximately the same way as an impulsive fireball. Figure 1 from Ramirez-Ruiz, Celotti, and Rees (2002; Figure 8.4) gives a schematic diagram illustrating the propagation of the jet and its cocoon through the stellar envelope beyond the core. In the case of a successful breakthrough of the jet these authors indicate that a strongly decelerated cocoon fireball could result. In this case, after the GRB and its afterglow, a rebrightening caused by the cocoon photospheric emission might appear with energy greater than $10^{50}$ erg. This component, essentially made of thermal UV and X-rays, is delayed by some seconds up to a day with respect to the GRB (see also Mészáros & Rees 2001). Other possibilities have been considered: for instance, if the relativistic jet carries less energy and inertia than the cocoon plasma, the collimated cocoon fireball can overtake the relativistic jet. In this situation the afterglow will be dominated by the emission of the cocoon material which is likely to be ejected at larger angles relative to the observer than the material of the jet itself (Ramirez-Ruiz, Celotti, & Rees 2002). These authors have also studied what happens when the cocoon becomes optically thin; shock waves within the plasma can contribute to a short-lived (few seconds) non-thermal $\gamma$/X-ray transient. Moreover, if magnetic dissipation within the plasma is important, it is also possible that a fraction of the energy stored in the cocoon contributes to a non-thermal UV/X-ray afterglow and excites Fe line emission from the envelope (see also Mészáros & Rees (2001) for similar conclusions).

Another interesting aspect of the jet breakout through the surface of the massive star is the behavior of the jet and its hot cocoon created as the jet traverses the star. At breakout the jet is tightly collimated by the pressure of the hot cocoon. After
breakout the cocoon pressure drops and the jet expands toward its natural opening angle. Lazzati and Begelman (2005) have shown that the evolving opening angle of the jet produces an angular stratification of the total energy. The angular structure is largely independent of the initial beam pattern. With the assumption of a constant luminosity they reproduce $\theta^{-2}$ profiles. Such an energy pattern $dE/d\Omega \propto \theta^{-2}$ was introduced by Rossi, Lazzati, and Rees (2002) and it produces a broken power-law afterglow (Rossi et al. 2004, Salmonson 2003, Granot & Kumar 2003). At the opposite of the uniform jet we get a structured jet which has been called USJ (see Section 6.7.4). The jet, the cocoon created during the sub-relativistic jet passage through the star, and the interactions between these two structures can be used to explain the complexity of temporal and spectral properties of the bursts, more complex than can be explained by the standard model.

In the next section we will discuss the (constraining) conditions required to have a ‘working’ collapsar, that is a collapsar which can produce GRBs. We discuss the types of stars, single or binaries, in which these conditions can be found. The discussion is mainly based on the work of Woosley and Heger (2006).
8.3 COLAPSARS, CONSTRAINTS ON THE PROGENITOR STAR

We have already presented the three basic requirements for the collapsar engine: (a) a black hole is necessary, (b) the hydrogen envelope around the collapsing star has to be ejected, preferentially in totality, allowing the jet to develop after breaking at the surface of the inner core, and (c) the presence of an unusually large angular momentum in the inner regions of the massive progenitor star is needed. This very fast rotation seems to be the key parameter that allows the formation of the relativistic jet.

The analysis of Woosley and Heger (2006) clearly shows that these criteria are not so easily fulfilled. One critical point is the large amount of angular momentum needed, which is one or two orders of magnitude greater than the angular momentum found in common pulsars (Woosley & Heger 2006). The discriminating characteristics of the stars that would produce GRBs are very likely their mass and their rotation rate. All currently favored GRB models require so much rotation that it also plays a decisive role in the explosion mechanism of the star. This is true for the collapsar model but, as we will see, this is also the case for millisecond magnetars (Usov 1992, 1994) and for the supranova model (Vietri & Stella 1998). To approach such extreme conditions, more massive stars, and especially those that have lost their hydrogen envelope early on, have to have cores which rotate more rapidly. Wolf–Rayet (WR) stars with a very rapidly rotating core are good candidates. But there is still the problem of magnetic torques, which are significant even when the envelope is removed. Magnetic torques coming from fields generated by the differential rotation have a tendency to enforce rigid rotation. Angular momentum is extracted from the inner core when it contracts and is transported to the outer layers which are spun up. Moreover, these magnetic torques brake the core when extensive mass loss slows down the rotation of the outer layers. For instance Heger, Woosley, and Spruit (2005) have found that magnetic torques decrease the final rotation rate of a collapsing iron core by about a factor of 30 to 50, when compared with the evolution of the same star without considering the effect of magnetic torques. Second, even without magnetic torques and even if the envelope of the star has been lost, the vigorous mass loss of typical WR stars still carries away a lot of angular momentum. There are two ways to solve this dilemma: the first assumes that a fraction of WR stars have a mass loss which is not standard, for instance with a decrease of up to a factor of 10 with respect to standard WR stars, while the second relies on the fact that long GRBs are produced at high redshift where the metallicity $Z$ is much lower than in the solar system. Under these conditions it is known (Vink & de Koter 2005) that WR stars have a lower mass loss rate. Another positive effect is expected in WR stars with low metallicity: they should rotate faster (Meynet & Maeder 2005). Finally, the mass loss anisotropies by stellar winds influence the loss of angular momentum (Maeder 2002). These anisotropies may result from mass-losses occurring predominantly at the poles. They can help in reducing the loss of angular momentum because the loss of polar mass takes away less angular momentum than isotropic mass-loss.

Woosley and Heger (2006) considered several models with bare He core or single stars of different mass and metallicity. They also considered a mass loss reduced by up to a factor of 10 from their standard values scaled by $\sqrt{Z}$. Their main results can be
summarized as follows (see also their Tables 1 and 2): all neutron stars resulting from stars that pass through a supergiant phase (red or blue) and do not lose their envelopes will rotate too slowly to become GRB progenitors. Even those stars resulting from He cores rotating near break-up (formed from binary or single stars) will be too slow to make GRBs unless their mass-loss rates are smaller than generally assumed (possibly due to low metallicity; Vink & de Koter 2005). So, how do we obtain GRBs? The progenitors might be the few O and B stars with high rotational velocity ($\sim 400 \text{ km s}^{-1}$) representing a few percent of this population. The condition for these stars is to retain enough angular momentum. Helium cores in binaries which have arrived at ignition with a rotation rate corresponding to $\sim 1/3$ of break-up are also good candidates if the number of such objects is sufficient (Woosley & Heger 2006).

Yoon and Langer (2005) are more optimistic. They find that the requirements of the collapsar model are rather easily fulfilled if the metallicity is sufficiently small. In this case rapidly rotating He stars (massive He stars appear in our Galaxy as Wolf–Rayet stars) are formed without the need to remove the hydrogen envelope, avoiding the spin-down induced by mass-loss. Such stars become He stars not by ejecting their H envelopes but by burning the H into He; the rapidly rotating core develops extensive mixing and nearly the entire H envelope can burn into He. Moreover, angular momentum transport from the He core to the H envelope by magnetic torques is insignificant. Their model suggests the existence of a lower mass limit of about 10 $M_\odot$ for the CO core for GRB production. This can be reached with an initial mass of only 20 $M_\odot$. Finally, the retention of sufficient angular momentum in CO cores in the range 10–40 $M_\odot$ (for initial masses of 20–60 $M_\odot$) allows the production of GRBs according to the collapsar scenario (Yoon & Langer 2005).

Fryer and Heger (2005) returned to the system consisting of a black hole surrounded by an accretion disk of high angular momentum material. They consider the merging of two He stars: the cores of massive binary stars would merge in a common envelope phase leaving a merged core and ejecting much of the hydrogen envelope without slowing down the rotation of the core (see their Figure 1 for the evolutionary path of close binaries), but in this scenario it is essential that the merged core rotates rapidly. Fryer and Heger (2005) found that in some cases the merger can produce cores that are rotating 3 to 10 times faster than single stars. Fryer et al. (1999) have considered three possible progenitors: single stars, the merger of a compact remnant and its binary companion (He merger), and the merger of two helium cores (He star/He star merger). Figure 8.1 provides the schematic evolution to the collapsar in these three cases. Single-star progenitors suffer from having insufficient angular momentum, which is not the case for He/He mergers. All three of these progenitors might be contributors to the GRBs population but their redshift distributions are very different.

Petrovic et al. (2005) have also followed the evolution of single stars and binary systems, with and without magnetic field. They show that the single-star model without magnetic field may retain enough core angular momentum to produce a GRB. The situation is quite different if magnetic fields are present, due to angular momentum transport by magnetic torques. In such a situation Heger et al. (2004) and Heger, Woosley, and Spruit (2005) concluded that the amount of angular momentum
in the core is one or two orders of magnitude lower than what is needed to produce a GRB. If the dynamo model of Spruit (2002) is correct, i.e. if it does not overestimate the magnetic torques, single massive stars cannot produce GRBs.

The evolution of close massive binaries was also considered by Petrovic et al. (2005). Here again the conclusion is that stellar cores with a specific angular momentum which is sufficient to produce a collapsar and a GRB can only be obtained with binary models without magnetic fields. But these conclusions are obtained for metallicities close to solar metallicity. Hirschi, Meynet, and Maeder (2005) have studied the evolution of massive single stars with various metallicities, taking into account rotation. They considered that WO Wolf–Rayet stars which have masses \( M > 50 M_\odot \) are possible GRB progenitors, these stars being only produced at low metallicities. But this study is done for single stars without magnetic field; this might be a severe limitation for the very fast rotation of the final black hole. Nevertheless the authors note that internal magnetic braking is most efficient during the red supergiant phase when the star has a fast rotating core and a slow-rotating envelope. If this phase is skipped, which is the case for the most massive WR stars \( (M > 60 M_\odot) \) the magnetic braking might be less effective. This result seems in contradiction with the work of Heger, Woosley, and Spruit (2005) who consider that single stars cannot retain enough angular momentum for GRB production, due to the strong effect of magnetic torques, concluding that binary stars are the most probable GRB progenitors. But it seems that Heger, Woosley, and Spruit (2005) studied massive star evolution only for solar metallicity; this might explain the differences in the conclusions.

Without being exhaustive we would like to conclude and to return to the three requirements for the collapsar engine, which have major consequences on the progenitor characteristics. We stress some interesting points which complete the previous discussion or result from it.

(a) The stellar mass above which a black hole can be formed is largely influenced by the rotation and the metallicity of the star. Fryer, Hungerford, and Young (2007) considered that the dividing line between neutron star and black-hole formation is located around 20 \( M_\odot \) but the fate of stars still depends sensitively upon stellar evolution codes. Since the collapsar model requires the formation of a black hole, it means that massive stars are involved. This seems to be a robust conclusion but it is difficult to specify the masses of these massive or very massive stars. This conclusion is supported by the fact that GRBs seem to occur most often in the brightest parts of their host galaxies, indicating that GRB progenitors have to arise from the most massive stars. Progenitor models also predict that it should be easier to form GRBs at lower metallicities, because low metallicities lead to weaker stellar winds, allowing the formation of more massive cores and black holes.

(b) The high angular momentum needed in the collapsing cores imposes strong constraints, making the loss of angular momentum through winds critical. How is the mass-loss kept sufficiently low to allow stars with very fast initial spin periods to retain enough angular momentum to produce an accretion disk? We have seen that low metallicity can reduce mass-losses. In binary progenitor
models the binary companion is used to remove the H envelope without losing angular momentum. Binaries might even be used to inject angular momentum (Fryer & Woosley 1998). Clearly, binary progenitors have fewer problems than single stars in achieving fast-rotating cores. This would be in agreement with the statement that a large fraction of massive stars (75% or more) are in close binary systems that will undergo mass transfer (Kobulnicky, Fryer, & Kiminki 2006), so that type Ib/c SNe would be primarily formed in binaries (van Dyk, Hamuy, & Filippenko 1996, Nomoto et al. 1995, Fryer, Hungerford, & Young 2007). Stellar evolution models (Heger et al. 2003) indicate that, at solar metallicity, single stars produce virtually no type Ib/c supernovae; single non-rotating stars can produce type Ib/c SNe at metallicity above 0.02 Z⊙, but these SNe which have weak shocks and eject little Ni will be dim. All these arguments concur to favor binary progenitors.

(c) Ejection of the H envelope: on the one hand we have seen that the stellar wind of single massive stars is a problem for the preservation of their angular momentum, but on the other hand the hydrogen envelope has to be removed if we want to see the jet. This dilemma can be avoided in the case of binary models, which invoke a phase of mass transfer that removes most or all of the H envelope. Another point concerning the envelope of GRB progenitors is the fact that GRB-associated SNe Ic have no H or He in their spectra, confirming that GRB progenitors are stars which have lost their H envelopes, but also part or all of their He envelopes (Fryer, Hungerford, & Young 2007).

The work of Fryer, Hungerford, and Young (2007) provides a detailed and critical analysis of collapsar progenitors. This is an interesting piece of work that gives, at the same time, an overview of the different models and of the observational status. They address the global properties of the GRB population, which can be used to discriminate among the various proposed progenitors: burst rates, type of associated SNe, metallicity, environment, morphology of the host galaxy. Their Tables 1 and 2 summarize the theoretical predictions of the models and the observational constraints. Among the collapsar progenitor models they discuss two broad classes. The first is the single-star progenitor with two variants: the classical single-star scenario with the loss of the H envelope through winds, and the mixing single-star scenario in which important mixing allows burning the hydrogen without needing to eject the hydrogen envelope. Metallicity requirements are severe for these single-star progenitors. The second class concerns binary progenitors. Among them are the merging of two massive stars both of which are burning He in their centres, with a phase of common envelope evolution, and the merger of a black hole with the He core of a massive star, or the merger of a black hole and a white dwarf.

Even though some types of progenitors seem to be favored, the above discussion clearly demonstrates that their nature is not at all settled: single stars, binaries with or without magnetic field, with solar or lower metallicities, are all possible. But it can be said without great risk of being contradicted that the progenitors represent a class of objects with exceptional properties: massive stars, certainly Wolf–Rayet, with very high rotational velocity, without H envelope and without important mass-loss,
massive stars in close binary systems being the favorite. The GRB rate would be only 0.2% of the SN rate (Woosley & Bloom 2006). Fryer, Hungerford, and Young (2007) estimated that about 1% of all core-collapse SNe (30% of them being of type Ib/c) would give GRBs, or that 1–10% of SNe Ib/c are associated with a GRB. This fraction coincides more or less with that of hypernovae: super-energetic SN explosions that exhibit higher energy and/or greater asymmetry than normal SN explosions. Finally, the production of GRBs is certainly favored in regions of low metallicity because their host galaxies and the regions where they are found have low metallicity (Vreeswijk et al. 2001, Savaglio, Fall, & Fiore 2003, Prochaska et al. 2004).

GRB progenitors would appear mainly in faint blue galaxies (Le Floc’h et al. 2003), with no long GRB associated with large spiral or with elliptical galaxies. GRB host galaxies are fainter and more irregular than the hosts of core-collapse supernovae (Fruchter et al. 2006). Their locations are strongly correlated with the blue light of their host galaxies (Bloom, Kulkarni, & Djorgovski 2002, Fruchter et al. 2006). In fact GRBs are more concentrated in the very brightest regions of their host galaxies than are core-collapse supernovae. This confirms that long GRBs occur preferentially in active star-forming galaxies, specifically in regions where the most massive stars die (see also Chapter 9).

Incidentally, and to conclude this section, we note that the problem of stellar rotation, which is so important for GRBs, is also crucial for the general issue of the evolution of massive stars, their collapse and explosion (Meynet & Maeder 2005). Fryer and Warren (2004), for instance, presented the results of 3D simulations of the collapse of rotating stars for a range of stellar progenitors. The rotation seems to have a major effect on the collapse through the torques exerted in differentially rotating regions by the magnetic fields that thread them (Heger, Woosley, & Pruit 2005). Moreover, the collapse calculated for spherically symmetric objects is not applicable because when the star is rotating the symmetry is broken, modifying the collapse and the explosion (Fryer & Warren 2004).

In the next section we will say a few words about a variant of the collapsar model, the supranova model of Vietri and Stella (1998), which is also associated with exploding massive stars. In this case the formation of a neutron star with a mass exceeding the critical mass is possible due to a very high rotation velocity. But after a time, which can be days to months, the loss of rotational energy leads to the collapse of the neutron star into a black hole and to the emission of a gamma-ray burst.

8.4 THE SUPRANOVA MODEL

As noted many times in Chapter 5, the fireball model requires a small baryon contamination (a few $10^{-4} M_\odot$ at most; Rees & Mészáros 1992). This reason pushed Vietri and Stella (1998) to consider that hypernova explosions (the collapsar model) may be problematic because the powerful winds of massive stars can significantly contaminate the environment in which the fireball develops. This is not the case for their supranova model.
Of course, mergers of two neutron stars or of a neutron star with a black hole are also able to produce a baryon-clean environment, but they have difficulties producing the majority of GRBs, the long ones. Another way of avoiding baryonic contamination is obtained if the collapsing core becomes a quark star as proposed by Paczyński and Haensel (2005). In this case a baryon–quark membrane appears at the surface of the collapsed compact object. This membrane can be penetrated by photons, electrons, positrons, neutrinos, and magnetic fields but not by baryons. This membrane prevents the baryon contamination of the energy outflow from a hot quark star.

In the model of Vietri and Stella (1998), the initial explosion is triggered by the gravitational implosion of a ‘hyper–massive neutron star’ (HMNS) into a black hole. Such stars rotate and have masses at infinity greater than the largest mass allowed for static compact stars. In their early evolution these HMNS lose angular momentum via the usual magnetic dipole radiation or gravitational waves. Slowly, the support of the centrifugal forces is lost, and the neutron star collapses into a black hole. Such an implosion is ‘silent’ with most matter being swallowed promptly into the hole. Since the neutrino mean free path inside the neutron star is expected to be small, the collapse will very likely be adiabatic. Given the high angular momentum of the HMNS, all the mass is not immediately accreted, and a torus of matter appears. The resulting configuration includes a black hole with a mass of 2–3 $M_\odot$, surrounded by a thin equatorial belt of matter with a mass estimated to be $\sim0.1 M_\odot$, sufficient to power a GRB (Vietri & Stella 1998). The extraction of energy from this system takes place via the conversion of the Poynting flux (due to the large-scale magnetic field locked into the torus of matter) into a magnetized relativistic wind (Usov 1992, 1994, Mészáros & Rees 1997, Katz 1997). The Blandford–Znajek mechanism (1997) can also be invoked to extract the much larger rotational energy of the black hole.

Even though the presence of the black hole with its torus of matter is reminiscent of the situation already found in the collapsar model, the situation is quite different here because the volume immediately surrounding the HMNS contains no baryons. The former SN explosion has swept away the medium surrounding the remnant. The small baryon contamination is a natural consequence of the supranova scenario, with typical values of a few $10^{-5} M_\odot$, well below the $\sim10^{-4} M_\odot$ upper limit required to get relativistic expansion with Lorentz factors of a few hundred. Moreover, unlike the collapsar model the GRB jet does not have to punch a hole through the stellar envelope.

The supranova scenario can also result when matter is slowly accreted by a neutron star from a surrounding accretion disk. This situation has been considered by Vietri and Stella (1999). In this case the formation of the HMNS is due to the transfer of mass and angular momentum from a close companion during a low-mass X-ray binary phase (LMXB). The loss of angular momentum through magnetic dipole radiation triggers the star’s collapse into a black hole. Of course in this case the time delay between the supernova and the collapse can be very large and the supernova remnant does not play any role in the GRB and its afterglow.

In these supranova scenarios, we can distinguish two steps. First, there is a specific supernova which leads to the formation of a HMNS. The second step, which occurs a few weeks to months later, leads to the collapse of the neutron star with the
production of a GRB. In the second scenario considered by Vietri and Stella (1999), involving an LMXB, these two steps can be largely separated in time. In the first scenario the neutron star formation implies young massive stars and, as for the collapsar model, GRBs will be associated with star-forming regions. In the second scenario the location of bursts inside their host galaxies will be the same as that of LMXBs, and GRBs are expected to cluster around galactic disks without any correlation with star-forming regions.

The detection of strong Fe emission features in the X-ray spectra of some GRB afterglows (Piro et al. 1999a,b, Yoshida et al. 2001), particularly that of GRB 991216 (Piro et al. 2000), renewed interest in the supranova model (see Section 3.7). Königl and Granot (2002), Inoue, Guetta, and Pacini (2003), and Guetta and Granot (2003) considered the situation before the HMNS collapses. The system involves a fast-rotating pulsar, which drives a powerful wind and a luminous plerionic nebula into the pre-burst environment. In this case the GRB should explode into a radiation-rich environment, i.e. into the luminous radiation field of the plerion. Some consequences for the ensuing GRB afterglow are analyzed by Inoue et al. (2003), and particularly the effect of inverse Compton cooling and the very high-energy emission in the GRB external shock. This emission is due to the plerionic radiation field, which acts as a seed for inverse Compton scattering, leading to the production of high-energy gamma-rays (GeV–TeV) in the afterglow. Other implications of this intense pulsar-type wind from the GRB progenitor have been analysed by Königl and Granot (2002).

Königl and Granot (2002) used the term pulsar wind bubbles (PWB) for the environment from which the afterglow emission originates, whereas Inoue, Guetta, and Pacini (2003) called it the plerionic nebula. There is a significant difference between the analyses of these two groups. Inoue et al. (2003) considered that the PWB interacts with the outlying SN ejecta material and that the Rayleigh–Taylor instabilities acting at the interface between the plerion and the SN ejecta interface induce filamentary clumping of the SN ejecta. For Königl and Granot (2002), on the other hand, the PWB remains largely confined within the SNR shell and the afterglow emission arises in the high-density interior gas, i.e. in a ‘plerionic’ SNR. The Crab and Vela remnants are examples of such ‘plerionic’ SNRs. Königl and Granot (2002) raised the problem of the values of \( \varepsilon_e, \varepsilon_B \) which must be high to achieve a high efficiency of the synchrotron afterglow emission (see Chapter 6). This is not so easy to obtain during the late phase of the afterglow in a typical ISM or stellar-wind environment. The presence of a PWB allows one to solve this difficulty because pulsar winds are expected to have a significant \( e^+ / e^- \) component and to be highly magnetized. They show that when the afterglow-emitting shock propagates inside a PWB, instead of inside a uniform ISM, the high electron and magnetic energy fractions (\( \varepsilon_e \) and \( \varepsilon_B \) ) inferred in a number of afterglow sources can be naturally accounted for. In summary, the supranova model can be used to explain some GRBs, with two advantages well outlined by the authors: the low baryonic contamination of the environment into which the GRB develops and the possibility of accounting for Fe-line X-ray emission features in the afterglow (of which we have few convincing examples though; Butler 2007).
We have seen the advantages of the collapsar and supranova models; as a conclusion it is interesting to indicate some of the open questions associated with these two models. These questions have been raised by Zhang and Mészáros (2004) in their review on gamma-ray bursts.

For the supranova we can summarize some of these questions:

- A fine-tuning is needed to make the GRB occur a few months to years after the formation of the neutron star (Shapiro 2000). In fact, Shapiro (2000) and Baumgarte, Shapiro, and Shibata (2000) noted that differentially rotating neutron stars can support significantly more rest mass than their non-rotating or uniformly rotating counterparts, but the stabilization due to differential rotation, although expected to last for many dynamical timescales (i.e. many milliseconds) will ultimately be destroyed by magnetic braking and/or viscosity. These processes drive the star into uniform rotation—which cannot support the excess mass and will lead to catastrophic collapse. So hypermassive stars can be produced but even if they last many dynamical timescales their life is very short in comparison with the timescales considered in the supranova model.

- Lacking an envelope the collapse may not produce a long burst (Böttcher 2001). Böttcher and Fryer (2001) investigated the prompt and delayed X-ray spectral features which can be observed in GRBs. They discussed previous calculations and in particular the use of the supranova model to explain the marginal detection of a redshifted Fe Kα emission line in the afterglow of GRB 970508 (Piro et al. 1999a). The authors noted that realistic calculations of collapsing neutron stars indicate that they eject too much baryonic material and have too little energy to produce GRBs (Ruffert, Janka, & Schaefer 1996, Fryer & Woosley 1998), but as these models did not include magnetic fields, they did not rule out the supranova model. However, such models would most likely produce short GRBs. Hence they have difficulties explaining the emission line feature seen in GRB970508 since that burst had a long duration with $t_\phi \sim 25$ s.

- As we will see, the association between SNs and GRBs is now well established, and the observations are compatible with the two events being quasi-simultaneous (e.g. Hjorth et al. 2003). Therefore the supranova model can only explain a minor fraction of observed GRBs, if any.

The collapsar is the leading scenario. We have seen its difficulties, a major one being the possibility of obtaining stellar progenitors with sufficiently high angular momentum. A second one is related to the wind environment, which is expected for GRB afterglows, and not commonly identified in afterglow observations (Frail et al. 2001, Panaitescu & Kumar 2001, and Sections 6.8 and 9.8.5). In fact, even though the inferred density varies significantly between the fits of different authors a constant density medium is consistent with the majority of afterglow data (see Zhang & Mészáros (2004) and the references therein for a review). A wind model works with
some bursts (see Chapter 6), but clearly they are not the majority, and we have indicated in Chapter 6 some suggestions to explain why this situation may exist (Chevalier, Li, & Fransson 2004, Scalo & Wheeler 2001). In spite of these difficulties we report in Section 8.6 several observations that strengthen the collapsar model. They include the detection of SN–GRB associations, the appearance of SN-like bumps in the optical afterglow light-curves, and other signatures of the presence of a supernova explosion or regions of massive star-formation.

In conclusion, the collapsar model has the best features today to explain long GRBs. Other models exist, however, based on similar progenitors (supernova giving birth to rapidly spinning magnetars) or, on the same final configuration, a massive black hole surrounded by a thick torus (mergers of compact objects). Models based on spinning magnetars are described in the following section, while the variety of merger models is described in Sections 8.7 and 8.8. Observational evidence linking long GRBs and supernovae is discussed in Section 8.6, and the more circumstantial evidence linking short GRBs with mergers of compact objects is presented in Section 8.10.

8.5 PULSAR MODELS

Several GRB models involve pulsar-like activity of the inner engine. In this case the outflow is not dominated by neutrinos which annihilate to produce $e^+/e^-$ pairs and radiation, but the outflow is magnetically driven or Poynting dominated. A Poynting flux can carry large amounts of energy through vacuum, providing a mechanism to transport energy without the need of matter. This is particularly interesting when we realize the difficulty that the models have in achieving a low baryonic load.

In Poynting flux powered outflows the rotational energy has to be transformed quickly enough into Poynting flux. For that to happen, magnetic fields have to be extremely high, of the order of $10^{15}$ G, about two orders of magnitude higher than in typical pulsars. These fields may be obtained at the formation of rapidly rotating neutron stars, for instance millisecond pulsars with extremely strong magnetic fields (Usov 1992). Duncan and Thompson (1992) have proposed that some pre-supernova stellar cores may have sufficient angular momentum to rotate near their break-up velocity, $P \sim 1$ ms, following collapse. Vigorous convection during the first few tens of seconds after the neutron star formation, associated with rapid rotation, makes the star a potential site for dynamo action. The resulting neutron star might generate a dipole field much stronger than $10^{13}$ G, thanks to an $\alpha-\Omega$ dynamo mechanism. In such stars the very fast differential rotation and the wrapping of field lines around the star by the shear motion allow the formation of magnetic structures of larger scale than in an ordinary pulsar. Such neutron stars can generate strongly magnetized relativistic winds. These neutron stars with very high magnetic field, called ‘magnetars’, have been invoked to explain the soft gamma repeaters (SGRs). Magnetars can
have dipolar magnetic fields in the range $10^{14}$–$10^{15}$ G, and they initially rotate with very short periods: $\sim$1 ms. Such neutron stars have a huge reservoir of magnetic energy.\(^1\)

Let us now go to the models in which the energy is extracted electromagnetically (see Usov, 1992, 1994, Thompson 1994, Mészáros & Rees 1997). Usov (1992), for instance, considered rapidly rotating neutron stars with surface magnetic fields $\sim$10$^{15}$ G. Such highly magnetized neutron stars might be formed by the collapse of accreting white dwarfs (WD) with anomalously high magnetic fields in binaries, like magnetic cataclysmic variables (but see Thompson (1994), who considers that such neutron stars cannot be millisecond pulsars). Being both rapidly rotating and highly magnetized, these neutron stars lose their rotational energy catastrophically on a timescale of seconds, or less. The rotation of the magnetic field creates a strong electric field. With such a strong electric field the vacuum is unstable (Schwinger 1951) and electron–positron pairs are created directly. Moreover, in very high magnetic fields photons can create pairs via the processes $\gamma + B \rightarrow e^+/e^-$ and $\gamma + \gamma' \rightarrow e^+/e^-$. These processes of $e^+/e^-$ pair production by electric fields are extremely powerful (Schwinger 1951). Usov (1992) shows that this pair plasma is initially optically thick and in quasi-thermodynamic equilibrium; it flows away from the neutron star at relativistic speeds. At a distance $r_{\text{rad}}$ from the pulsar, where the optical depth for X-rays and $\gamma$-rays is $\sim$1, the radiation propagates freely. At this distance $r = r_{\text{rad}} \sim 10^8$–$10^9$ cm, the mean Lorentz factor of the plasma particles can reach a few times $10^2$. The typical energy of the $\gamma$-rays radiated by the $\gamma$-ray photosphere of the $e^+/e^-$ wind at $r = r_{\text{rad}}$ is a few times (0.1–1) MeV (Prilutski & Usov 1975, Horstman & Cavallo 1983), with a high energy tail which can extend to a few hundred megaelectronvolts. With this model Usov explained both short and long GRBs and their main characteristics (see also Usov 1994). Here the radiation of the pulsar consists of two components. One is thermal radiation with a black-body-like spectrum, and an equivalent temperature $\sim$$10^{10}$ K ($\sim$1 MeV). This radiation is emitted by the $\gamma$-ray photosphere of the outflowing $e^+/e^-$ plasma at the distance already defined $r_{\text{rad}} \sim 10^8$ cm. To have the non-thermal component, a radius of $10^{13}$ cm has to be reached. This is the limit where the magnetic field is frozen into the outflowing plasma; the MHD approximation for the pulsar wind breaks down and intense electromagnetic waves are generated, with frequencies equal to the frequency of the pulsar rotation. At the wind front, the outflowing particles are accelerated in the field of these intense electromagnetic waves to very high Lorentz factors ($\sim$$10^6$). By interaction with the fields of the electromagnetic waves they generate a non-thermal radiation called synchro-Compton radiation. The typical energy of the

\(^1\)Jet-like Poynting flows from a torus around a black hole have also been considered. These systems can produce intense GRBs requiring efficiencies which can be relatively low (Mészáros & Rees 1997). In this case the torus around the black hole may be highly magnetized due to the field amplification by differential rotation (Narayan, Paczyński, & Piran 1992, Thompson 1994, Mészáros & Rees 1997, Katz 1997). In these systems a pulsar-like mechanism is at work, the GRB being produced by electrodynamic processes that are able to extract energy from accretion disks and to transform the rotational energy into particle acceleration as in pulsars.
non-thermal component is $\sim 1$ MeV. A high-energy tail to $10^4$ MeV may exist. The baryonic matter which can be ejected occasionally as a result of some kind of plasma instability can suppress the $\gamma$-ray emission episodically possibly leading to the time structure of the GRBs (Usov 1994).

Smolsky and Usov (2000) considered the interaction of a strongly magnetized relativistic pulsar wind with the ambient medium and the resulting non-thermal emission of the wind. This wind is always assumed to come from a fast-rotating compact object like a millisecond pulsar with a surface magnetic field of $10^{15} - 10^{16}$ G. Usov (1994) identified a first radiating region at $r \sim 10^{13}$ cm emitting powerful non-thermal emission. This radiation was assumed to have been produced by electrons accelerated in the fields of LAEMW (large-amplitude electromagnetic waves) which are generated by the oscillating currents at the wind front (see Smolsky & Usov 2000). The introduction of LAEMW represents a way to extract rotational energy and to transform it into magnetic field energy, giving a way to avoid internal processes involving shocks. Smolsky and Usov (2000) introduced a second external radiating region at a distance $r \sim 10^{16} - 10^{17}$ cm, which is the radius where the deceleration of the wind due to its interaction with the ambient medium becomes significant. They confirmed the result of Mészáros and Rees (1992) that the interaction of the wind with the ambient medium may be responsible for X-ray and $\gamma$-ray emissions. In the case of Smolsky and Usov (2000) the interaction of the wind with the ambient medium differs from the conventional model involving collisionless shocks. In their approach collisionless shocks are replaced by a mechanism of electromagnetic acceleration at the wind front. Large amplitude electromagnetic waves are embedded in a region ahead of the wind front and the electrons highly accelerated in the fields generated by these LAEMWs emit synchro-Compton radiation, which closely resemble synchrotron radiation. As noted by Blackman and Yi (1998), LAEMW emission mechanisms do not require the extremely rapid magnetic field amplification, unlike blastwave models.

Duncan and Thompson (1992) also discussed the case of neutron stars with unusually strong magnetic fields: the magnetars. Young magnetars ($10^4$ yr old) are proposed as the sources of the soft gamma repeaters (SGRs). As these stars contain a tremendous reservoir of magnetic energy they proposed that GRBs could be triggered by magnetic reconnection, similar to stellar flares, in the magnetospheres of magnetars (Sturrock 1986). It is known that flares are triggered in magnetically active main sequence stars when convective motions displace the footpoints of the field sufficiently to create tangential discontinuities which undergo catastrophic reconnection. Duncan and Thompson (1992) considered the possibility that similar reconnection events occur in magnetars where the footpoint motions can be driven by the diffusive processes they analysed. Sturrock (1986) had previously proposed flares in neutron star magnetospheres as the basic energy release mechanism for GRBs. Reconnection events lead to electric fields parallel to the magnetic field. These electric fields accelerate electrons which produce high-energy gamma-rays which annihilate in the strong magnetic field, leading to an electron–positron cascade. The $e^+ / e^-$ can radiate by the synchrotron process. A possible contribution due to curvature radiation might explain the high-energy tail observed in some GRB spectra. The scenario proposed
by Duncan and Thompson (1992) was studied thoroughly by Thompson (1994), starting with a MHD wind which is relativistic and Poynting-flux-dominated. This wind may emanate from a rapidly rotating neutron star, or a neutron disk in which a poloidal field \( >10^{14} \) G has been generated by a helical dynamo. Such objects could be produced by failed type Ib supernova, by accretion-induced collapse of a WD, or perhaps by a binary neutron star merger. Mildly relativistic Alfvén turbulence is excited in the wind by reconnection or by hydrodynamical instabilities. Alfvén turbulence can introduce random electron motions and its energy is transferred very effectively to the electrons (Thompson 1994). Gamma-rays are generated via comptonization; i.e Compton upscattering by mildly relativistic Alfvén turbulence of quasi-thermal radiation in ultra-luminous and highly compact relativistic winds. This model requires converting a significant fraction of the Poynting flux into Alfvén turbulence and thence into gamma-rays (see also Thompson & Duncan (1993) for other mechanisms that can convert the magnetic energy generated inside such stars into gamma-rays, the question always being to avoid significant amount of baryonic contamination). With this model the \( \gamma \)-ray spectrum is generated at a very small distance \( \sim 10^9 \) cm, from the source, without the need for any interaction with an external medium. Thompson also noticed that in this model the strong small-scale magnetic fields generated by the convective motions can trap most of the baryons near the stellar surface, thereby suppressing the steady mass-loss rate by a large factor (10 or more).

Drenkhahn and Spruit (2002) discussed various aspects of the conversion of Poynting flux energy into gamma-rays. They investigated the effects of magnetic energy release by local magnetic dissipation processes in Poynting-flux-powered GRBs. As any asymmetry of the magnetic field leads to small-scale (wave-like) variations in the electromagnetic field carrying the energy outward, the authors assumed that these small-scale irregularities are subject to rapid reconnection. Thus the magnetic field can release its free energy stored in the small-scale field variations. This energy can be converted into radiation. The magnetic field acts as an energy reservoir. A significant part of the Poynting flux is converted into kinetic energy, and the other part is converted into heat. Assuming that the reconnection processes under optically thin conditions maintain a significant population of energetic electrons, these electrons radiate their energy through synchrotron radiation as in the standard model. Here, of course, the magnetic field is a natural part of the flow itself and radiation results from the local dissipation of magnetic energy and not from shock conversion of kinetic energy (Drenkhahn & Spruit 2002). These authors also considered the case of higher magnetic energy into radiation—a scenario that might explain X-ray flashes.

Wheeler et al. (2000) also considered the case of a newly born neutron star resulting from the core collapse of a massive star. This collapse might be strongly asymmetric. If the neutron star is rotating and magnetized when it forms, the cooling phase will be associated with contraction, spin-up, and amplification of the magnetic field. The authors consider that GRBs stimulated by Poynting flux from the neutron star will occur near the end of the cooling and contraction stage, when the neutron star spins with a period \( \sim 1 \) ms. This appears as a necessary condition to power a
GRB, even if it is not sufficient. The amplification of the magnetic field leads to a final, amplified field which might be around $10^{16}$ G. So the context of a magnetar is also present here. Differing from the previous authors, no Alfvén waves are considered; rather they involve the generation of ultra-relativistic MHD waves (UMHDW) where the density is high, and large amplitude electromagnetic waves (LAEMW) where the density is low. The production of traditional Alfvén waves as an energy-loss mechanism is neglected by Wheeler et al. (2000). They argue that Alfvén waves occur when the magnetic field is relatively weak and the waves are a propagating perturbation. Alfvén waves will thus be generated beyond the magnetosphere where the magnetic field no longer dominates the particle dynamics. In the proto-neutron star case, Alfvén waves could be generated in the region beyond the magnetosphere and within the extended light-cylinder. Once the neutron star contracts and the light-cylinder passes within the magnetosphere, contracting from beyond $R_A$ to significantly less than $R_A$ where Alfvén waves are not a relevant concept, the structure of any organized dipole field is disrupted and the dominant losses will occur at the light-cylinder by generation of UMHDW. The production of Alfvén waves and associated losses thus seem to be a secondary process (Wheeler et al. 2000).

Tight collimation of the original matter jet (formed promptly during the proto-neutron star phase) and of the subsequent flow of UMHDW in a radiation-dominated jet is expected. If a UMHDW jet is formed it can drive shocks propagating along the axis of the initial matter jet. The shocks associated with the UMHDW jet could generate gamma-rays by the Colgate mechanism (Colgate 1968, 1973, 1974) as they propagate down the density gradient at the tip of the jet. Alternatively, bulk proton acceleration could bring them above the pion production threshold. Pions, which are copiously produced through collisions with the surrounding wind, trigger a cascade of pairs and $\gamma$-rays from high to low energy. The authors have also suggested that, if the environment where UMHDW are propagating reaches a low density, LAEMW and pair cascades can be induced (see also Thompson & Madau 2000). These different waves are considered good candidates to produce a collimated GRB. However, the authors stress that the outlined processes may not be as effective as they assumed (see the discussion in the conclusion of Wheeler et al. (2000)). In any case rotation and magnetic field have a strong potential to create axial-matter-dominated jets that will drive strongly asymmetric explosions for which there is already ample observational evidence in type II and type Ib/c supernovae. Strong flows of UMHDW/LAEMW waves serve to reinforce the possibility of generating asymmetric explosions (Wheeler et al. 2000).

Independent of the GRB question, the problem of the birth and the fate of these highly magnetized stars is of great importance by itself. Even if magnetars are rare, they might represent 10% of pulsar births. We have seen two possible sources of relativistic Poynting-dominated winds: millisecond pulsars with high magnetic fields or with a highly magnetized torus. In all cases the rotational energy of a compact

$^2$ $R_A$ is the Alfvén radius defined as the distance at which the ram pressure balances the magnetic pressure.
object is tapped and the rotating magnetic field produces a Poynting flux. We have also seen different ways of transforming the Poynting flux into a GRB. The energy can be extracted electrodynamically from the accretion disk, but other ways have been explored: the formation of strong electric fields near the rotating magnetic neutron star; the generation of Alfvén turbulence excited by reconnections; the production of ultra-relativistic MHD waves and of large-amplitude electromagnetic waves. After this brief excursion into the realm of fast-rotating highly magnetized neutron stars let us go back to the more widely discussed collapsar model and to the observational evidence that supports it.

8.6 OBSERVATIONAL EVIDENCE LINKING GRBS WITH THE COLLAPSE OF MASSIVE STARS

During the last few years our understanding of type Ib/c supernovae and their connections with long-duration GRBs has been revolutionized. Excellent reviews of our current understanding of the GRB–SN association have been given by Soderberg (2006), and Woosley and Bloom (2006), with a comprehensive list of references. In this section, we analyze the observations that have consolidated these associations, without, however, proving that all long soft GRBs are accompanied by type Ic supernovae.

8.6.1 GRB–SN association

We discuss here the first GRBs which have been unambiguously associated with SNe. Table 8.1 (from Woosley & Bloom 2006) provides a summary of convincing GRB–SN connections up to 2005.

- The first case of a GRB–SN connection was GRB 980425, which has been associated with SN 1998bw, a remarkable nearby and very bright type Ic supernova with unusual radio, optical, and spectroscopic properties. The very broad optical spectral features which were observed in SN 1998bw required a large explosion energy, more than 10 times that of an ordinary supernova; this is why it was termed a ‘hypernova’. This SN is located inside the error box of the GRB and, at the time of its discovery, the chance of coincidence was estimated to be \( \sim 10^{-4} \). GRB 980425 is the closest burst recorded to date, with redshift \( z = 0.0085 \) (about 37 Mpc, Galama et al. 1998a,b). The rapid rise of the radio emission from the SN showed that the SN explosion was simultaneous with the GRB to within a day. The energy of the GRB in \( \gamma \)-rays was \( 8.5 \pm 0.1 \times 10^{47} \) erg, assuming isotropic emission. This is more than three orders of magnitude lower than the majority of long-duration GRBs (Frail et al. 2001, Bloom, Frail, & Kulkarni 2003).

- The second association was GRB 030329 and the supernova SN 2003dh (Hjorth et al. 2003, Stanek et al. 2003; see also Section 4.3). This GRB, detected by HETE-2, was very bright (with a fluence in the top 0.2% of the 2704 GRBs detected by BATSE) and lasted about 25 s. A high-resolution VLT spectrum gave its redshift, again low: \( z = 0.1685 \) (Greiner et al. 2003b). The optical afterglow
Table 8.1. GRB–SN associations up to mid-2005 (from Woosley & Bloom 2006).

<table>
<thead>
<tr>
<th>Name of burst/SN</th>
<th>$z$</th>
<th>$Peak_{V}$ (mag)</th>
<th>$T_{peak}$ (day)</th>
<th>SN likeness/description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRB 980425/1998bw</td>
<td>0.0085</td>
<td>$M_V = -19.16 \pm 0.05$</td>
<td>17</td>
<td>Ic-BL</td>
<td>Galama et al. 1998b</td>
</tr>
<tr>
<td>GRB 030329/2003dh</td>
<td>0.1685</td>
<td>$M_V = -18.8$ to $-19.6$</td>
<td>10–13</td>
<td>Ic-BL</td>
<td>Hjorth et al. 2003; Stanek et al. 2003; Bloom et al. 2004; Lipkin et al. 2004</td>
</tr>
<tr>
<td>XRF 020903</td>
<td>0.25</td>
<td>$M_V = -18.6 \pm 0.5$</td>
<td>$\sim 15$</td>
<td>Ic-BL</td>
<td>Soderberg et al. 2005</td>
</tr>
<tr>
<td>GRB 011121/2001dk</td>
<td>0.365</td>
<td>$M_V = -18.5$ to $-19.6$</td>
<td>12–14</td>
<td>I (IIIn?)</td>
<td>Bloom et al. 2002; Garnavich et al. 2003; Greiner et al. 2003a</td>
</tr>
<tr>
<td>GRB 050525A</td>
<td>0.606</td>
<td>$M_V \approx -18.8$</td>
<td>12</td>
<td>I</td>
<td>Della Valle et al. 2006</td>
</tr>
<tr>
<td>GRB 021211/2002lt</td>
<td>1.00</td>
<td>$M_V = -18.4$ to $-19.2$</td>
<td>$\sim 14$</td>
<td>Ic</td>
<td>Della Valle et al. 2004</td>
</tr>
<tr>
<td>GRB 970228</td>
<td>0.695</td>
<td>$M_V \sim -19.2$</td>
<td>$\sim 17$</td>
<td>I</td>
<td>Galama et al. 2000; Reichart 1999</td>
</tr>
<tr>
<td>XRR 041006</td>
<td>0.716</td>
<td>$M_V = -18.8$ to $-19.5$</td>
<td>16–20</td>
<td>I</td>
<td>Stanek et al. 2005; Soderberg et al. 2006c</td>
</tr>
<tr>
<td>XRR 040924</td>
<td>0.859</td>
<td>$M_V \sim -17.6$</td>
<td>$\sim 11$</td>
<td>?</td>
<td>Soderberg et al. 2006c</td>
</tr>
<tr>
<td>GRB 020405</td>
<td>0.695</td>
<td>$M_V \sim -18.7$</td>
<td>$\sim 17$</td>
<td>I</td>
<td>Price et al. 2003b</td>
</tr>
</tbody>
</table>

$^a$ The time of peak brightness is reported in the rest-frame if the redshift is known, in the observed frame otherwise.
was unusually bright with a magnitude of 12.6 in the R band 1.5 h after the burst, but it was also very bright at other wavelengths: radio (Berger et al. 2003), millimeter (Kuno, Sato, & Nakanishi 2003), submillimeter (Hoge et al. 2003), and X-rays (Marshall & Swank 2003). For this burst Berger et al. (2003) proposed a two-component explosion model with a narrow jet (∼5°) and a wider jet (∼17°), this second component powering the radio afterglow and the late optical emission (see Section 6.7.4 on structured jets). In fact this burst was the first for which it was possible to obtain a spectroscopic signature for an underlying supernova identified in the optical afterglow (Stanek et al. 2003). Spectra obtained a few days after the GRB (on April 5, 2003) showed broad peaks characteristic of energetic supernovae (see Figure 4.5). The spectrum of the SN is remarkably similar to the type Ic ‘hypernova’ SN 1998bw associated with GRB 980425 (see Figure 4.5). In addition, Lipkin et al. (2004) using a very complete compilation of BVRI light-curves, which involved five observatories, concluded that the evolution of SN 2003dh was well described by a light-curve similar to that of SN 1998bw but fainter by 0.3 magnitudes and with a timescale stretched by a factor of 0.8. The analysis of this SN was complicated by the brightness of the optical afterglow, which outshone that of the supernova for the first week and was still competing with it for at least one month. For SN 1998bw this problem did not exist because no optical afterglow was detected. For SN 2003dh, the study of the optical curves covering about 6 days before supernova maximum to about 60 days after allowed Deng et al. (2005) to suggest that the progenitor was a main sequence star with a mass of ∼25–40 $M_\odot$.

- The third well-established GRB–SN association was GRB 031203 at $z = 0.1055$ and the bright and highly reddened supernova SN 2003lw (Bailyn et al. 2003, Bersier et al. 2004, Thomsen et al. 2004, Cobb et al. 2004). This GRB, which was discovered by INTEGRAL, was a faint burst with an isotropic energy smaller than 10$^{50}$ erg (Sazonov, Lutovinov, & Sunyaev 2004) and a duration ∼30 s (Mereghetti & Gotz, 2003). The afterglow was faint with a rebrightening similar to the light-curve of SN 1998bw if one assumes that the GRB and the SN explosion were almost simultaneous. Mazzali et al. (2006b), accounting for the reddening, found that SN 2003lw was 0.3 mag brighter than the prototypical GRB–SN 1998bw. Spectra taken close to the maximum of the rebrightening showed extremely broad features, like SN 1998bw (Malesani et al. 2004). Modeling the light-curve and the spectrum of the SN simultaneously, Mazzali et al. (2006b) found that the optimal explosion could be modeled with an energy of ∼6 × 10$^{52}$ erg and an ejected mass $M_{ej} ∼ 13 M_\odot$, larger than for SN 1998bw. From $M_{ej}$ the inferred mass of the progenitor is 40–50 $M_\odot$ (Mazzali et al. 2006b). For these authors the similarity between these two GRBs and their relative faintness suggests that both GRBs may be normal events viewed slightly off-axis or a weaker but possibly more frequent type of GRB. Daigne and Mochkovitch (2007) showed that the second scenario appears to be more realistic and that the rate of weak events like GRB 980425 in the Universe has to be much higher than the rate of ‘standard’ bright gamma-ray bursts. Based on radio calorimetry and on the absence of the signature of an off-axis afterglow, Sazonov, Lutovinov, and
Sunyaev (2004) and Soderberg et al. (2004) have suggested that GRB 031203 was intrinsically under-luminous. This event also gave the opportunity to observe a very nice expanding X-ray halo with XMM (see Figure 4.7; Vaughan et al. 2004). Small-angle scattering of X-rays by dust grains can produce X-ray halos around distant X-ray sources (Overbeek 1965). Vaughan et al. (2004) attributed the rings to galactic dust concentrated in two distinct slabs located at distances around 900 and 1400 pc.

In conclusion, it is interesting to note that the three SNe involved in convincing associations with GRBs have relatively similar spectra and light-curves (but see Gal-Yam et al. 2004 and Cobb et al. 2004), independent of the fact that the three GRBs are very different and that hypernovae and SN Ic present in general a large heterogeneity (Deng et al. 2005).

To this short list of SN–GRB associations we must add the remarkable association of GRB 060218 with SN 2006aj (Campana et al. 2006a). This discovery was discussed extensively in Section 4.10. This association is exceptional in several respects and it has been considered to be a major discovery of Swift. The burst detected by the BAT instrument onboard Swift was very long in $\gamma$-rays (2100 $\pm$ 100 s). Its fluence was dominated by soft X-ray photons and this burst can be classified as an X-ray flash. Soon after the discovery, low-resolution spectroscopy of the optical afterglow and of the host galaxy revealed strong emission lines at redshift $z = 0.033$ ($\sim$145 Mpc). The isotropic equivalent energy of the prompt GRB was a few times $10^{49}$ erg, making GRB 060218 an under-luminous GRB. This is two orders of magnitude fainter than GRB 030329 associated with SN 2003dh at $z = 0.169$, but two orders of magnitude brighter than GRB 980425 associated with SN 1998bw at $z = 0.0085$. The presence of a rising SN was found three days after the burst (Masetti et al. 2006, Fugazza et al. 2006, Soderberg et al. 2006b, Pian et al. 2006). The broad emission features observed in the optical spectrum were consistent with a type Ic SN with a lack of hydrogen. This SN was designated SN 2006aj. Its spectrum can be compared with other broad-line SNe without GRBs and with SN 1998bw (Modjaz et al. 2006; Figure 4.17). Pian et al. (2006) compared its light-curve with those of the other three GRB–SNe, and with two broad-line SNe and one normal SN. The maximum luminosity of the four GRB–SNe spans a much narrower range than the maximum luminosity of the entire class of SNe Ic. The rapid rise to maximum of SN 2006aj (Figure 8.5) and the evolution of the photospheric expansion velocity are suggestive of a highly asymmetric explosion (see also Mazzali et al. 2006b).

The most striking feature of this burst is the presence of a soft component appearing in the X-ray spectrum from 159 s up to $\sim 10^4$ s. This component can be fitted with a black body whose temperature decreases marginally between $t = 300$ s and $t = 2600$ s, corresponding to an emission region that expands and shifts into the optical UV band as time passes (Campana et al. 2006a; Figure 4.18). This component has been interpreted as shock break-out from a dense wind surrounding the SN progenitor (Campana et al. 2006a). If this interpretation is correct it would be the first direct measurement of the shock break-out of the stellar envelope and the stellar
wind (see the first investigation of this question by Colgate 1968, 1974). Moreover, from their analysis Campana et al. (2006a) concluded that the progenitor of this XRF was a compact massive star, most probably a WR star. This is consistent with the idea that GRBs and XRFs share a common origin: massive WRs. Finally the quasi-simultaneity of the non-thermal (GRB) and thermal X-ray emissions seems to indicate that the SN explosion and the XRF occurred very close in time.

Other XRFs have been associated with supernovae, like XRF 030723 (Fynbo et al. 2004) or XRF 020903 (Soderberg et al. 2005). But for many others, such as XRF 040701 at $z = 0.21$, XRF 040812 or XRF 040916 (at unknown distance) studied by Soderberg et al. (2005) there is no evidence for associated SNe. These authors concluded that XRF–SNe exist but can be significantly fainter than SN 1998bw. This might suggest the existence of a large population of such low-energy events, which could be more abundant than common classical GRBs, perhaps the most abundant form of high-energy transients of stellar origin in the universe (Pian et al. 2006). Due to the limited sensitivity of present-day instruments, only the nearest/brightest members of this population would be detected (see also Section 4.5.3).
Returning to the subject of GRB 060218, its association with a SN and the detailed afterglow studies that followed seemed to confirm the collapsar model, at least for a fraction of long GRBs. Nevertheless, the analysis of Fan et al. (2006) indicated that for various reasons this event appeared to be incompatible with the classical collapsar model. The lack of a ‘jet break’ up to 10^6 s after the burst indicated an outflow that was quite wide (θj ≥ 50°), for the usual collapsar model in which a narrow jet has to punch a narrow hole in the envelope of the progenitor (but see the discussion on jet breaks in Chapter 7). Moreover, the initial Lorentz factor was very low (this can explain the unusual soft spectrum) and the medium surrounding the progenitor had an ISM low-density profile, in contrast with the expected dense stellar wind (but this situation has been observed for many long GRBs; see Chapter 6). This could be due to the progenitor’s wind which was weak during the many hundred years before the burst. Fan et al. (2006) concluded that GRB 060218 was almost a ‘failed GRB’, with a relativistic ejecta unable to make its way through the envelope, even though the kinetic energy of the blast wave was rather high: 5 × 10^{50} erg. These authors also considered unlikely alternative explanations of this peculiar GRB based on viewing angle effects.

We will see at the end of this Section 8.6 that two nearby long bursts, GRB 060505 and GRB 060614 are not accompanied by supernova emission, showing that the GRB–SN association is probably not always present. Let us now review other signatures of this association, which are less remarkable and do not allow such clear-cut conclusions.

8.6.2 Bumps in optical afterglow light-curves

One example is the association of GRB 011121 with SN 2001ke (see Figure 3.12). The GRB was a bright burst detected by BeppoSAX (Piro et al. 2001). Spectra of the optical afterglow, obtained with the Magellan 6.5-m Telescope, gave a redshift z ∼ 0.36 (Garnavich et al. 2003, Infante et al. 2001). Late-time imaging and spectra, taken 2 weeks after the GRB, suggested a slowdown of the initial fast decline of the optical afterglow. Hubble Space Telescope images taken 2 weeks after the burst were analysed by Garnavich et al. (2003) who found that the source was two magnitudes brighter than the extrapolation of the optical afterglow and that the colors were inconsistent with a power-law spectrum, suggesting the presence of a supernova, called SN 2001ke (Stanek et al. 2002). This supernova was bluer near maximum than SN 1998bw and faded more quickly. This was the first spectrum ever obtained of a GRB supernova at cosmological distance (taking into account the fact that SN 1998bw was associated with a very unusual GRB). The blue color was consistent with a supernova interacting with circumstellar gas. The light-curve of SN1998bw is a poor match to SN 2001ke (Garnavich et al. 2003). Greiner et al. (2003a) found this association convincing because three signatures (jet break, wind density profile, and SN bump) were found.

There are other GRBs whose late-time optical afterglows suggest the appearance of a SN, which is more or less visible as a late optical bump lasting several days. Such SN contributions where reported for GRB 980326 (Castro-Tirado & Gorosabel 1999,
Bloom et al. 1999), GRB 970228 (Galama et al. 2000), GRB 970508 (Sokolov 2001), GRB 020405 (Price et al. 2003b), and possibly for GRB 980703 (Holland et al. 2001), GRB 970712 ( Björnsson et al. 2001), GRB 991208 ( Castro-Tirado et al. 2001), and GRB 000911 ( Lazzati et al. 2001).

An interesting example is the discovery of the optical transient associated with GRB 020410. The transient was found by HST thanks to observations taken 28 and 65 days after the burst. This was the first time that the optical afterglow of a GRB was discovered by an orbiting observatory (Levan et al. 2005). The late-time HST data taken ~100 days after burst for comparison implied a very red transient. The optical transient brightening was estimated between day 7 and day 28. The data can be explained by the presence of a type Ib/c SN at a redshift \( z \sim 0.5 \), but it seems that it would be the faintest SN (fainter than SN 1998bw by two magnitudes) yet associated with a GRB (Levan et al. 2005). Therefore the HST observations correspond to the detection of the SN without the afterglow (a deep non-detection of the afterglow was reported one week after the GRB). This event was interesting because it was the first identification of an optical counterpart of a GRB thanks to the supernova light. It also demonstrates that without the deep late HST observation this burst would have been classified as a dark GRB. In addition it indicates that it is quite possible that similar SNe could be present in the afterglows of many long GRBs (Levan et al. 2005).

A systematic analysis of SN light-curves in GRB afterglows was done by Zeh, Klose, and Hartmann (2004) for 21 bursts from 1997 to 2002 with established redshifts. The light-curves were fitted with the light from the afterglow, from the host galaxy and from a SN component scaled to SN 1998bw. The most interesting result of this analysis was the evidence for a late-type bump in all GRB afterglows with measured redshifts smaller than \( z \sim 0.7 \). At higher redshift, possible SN bumps were confirmed for GRB 980703, GRB 000911, and GRB 021211. For five other bursts (GRB 970508, GRB 991216, GRB 000418, GRB 010222, and GRB 020813) with \( 0.7 < z < 1.5 \) only upper limits on the luminosity of an underlying SN component were established. The conclusion of the authors was that their work supported the idea that all afterglows of long-duration GRBs contain light from an associated supernova, SN 1998bw being at the bright end of the SN luminosity distribution (Zeh, Klose, & Hartmann 2004).

Another analysis (Soderberg et al. 2006c) concentrated on HST observations of SNe accompanying two GRBs: GRB 040924 and GRB 041006. Both GRBs were discovered by HETE-2 and their redshifts are respectively \( z = 0.859 \) and \( z = 0.716 \). These bursts can be associated with supernovae like SN 1998bw, respectively dimmed by \( \sim 1.5 \) and \( \sim 0.3 \) magnitudes (see for example the very convincing fit of the light-curve of the afterglow of GRB 041006 with SN 1998bw at \( z = 0.86 \) dimmed by 0.3 mag; Figure 5 of Soderberg et al. 2006c). Deep photometry of the afterglow of GRB 041006 (Stanek et al. 2005) led to similar conclusions: the existence of a bump well fitted by the light-curve of SN 1998bw at \( z = 0.716 \) dimmed by 0.3 magnitude, with a stretching factor of 1.38 in time (Figure 8.6). For these two events there is thus a significant spread in the luminosities of their associated supernovae.
Figure 8.6. Late R band light-curve of the afterglow of GRB 041006. The early power-law decay of the optical afterglow (a $t^{-1.3}$ decay) is shown with the dashed line. A SN 1998bw light-curve extended by a factor 1.38 is indicated by the dotted curve. The solid line shows the power-law combined with the stretched SN light-curve. This is a good example of a light-curve “bump”, which strongly suggests the late-time dominance of an underlying SN component (from Stanek et al. 2005).

To better study the dispersion in luminosity of GRB-associated SNe, Soderberg et al. (2006c) compiled the peak optical magnitude and $^{56}$Ni mass estimates of GRBs and XRFs associated with SNe (Figure 8.7). A comparison is made with local SNe Ibc. The redshifts of XRFs and GRBs in their study range from 0.0085 to 1.058. The upper panel of Figure 8.7 displays the $^{56}$Ni mass estimate versus the peak optical magnitudes for GRB/XRF associated SNe and for local SNe Ibc. This figure shows that GRB-associated SNe do not necessarily synthesize more $^{56}$Ni than ordinary SNe Ibc and that the peak magnitude scales roughly with $^{56}$Ni mass for both populations. The lower panel of Figure 8.7 compares the peak optical magnitude for the two SN populations. The two samples exhibit noticeable overlap, with GRB-associated SNe clustering toward the brighter end of the distribution. Nevertheless, some of the faintest SNe Ibc observed are in fact associated with GRB and XRFs. Clearly this comparison shows that GRB/XRF associated SNe are not always over-luminous in comparison with local SNe Ibc (see also Matheson et al. 2001 for a study of many optical spectra of SNe Ibc). The same conclusion is reached by Woosley and Bloom (2006). In their Figure 7 they compare secure SN bumps found in GRB afterglows with the peak absolute $M_v$ of the local SNe Ibc population. At first glance this comparison shows that GRB–SNe populate the bright subclass of the Richardson, Branch, and Baron (2006) sample. However, when non-detections are folded in, Woosley & Bloom (2006), concluded that the GRB–SNe population is statistically
consistent with having been drawn from the local SN Ibc population studied by Richardson, Branch, and Baron (2006), and the connection with SNe Ic is favored. This may imply that both populations share a common progenitor system and/or explosion mechanism (Soderberg et al. 2006c).

A last point concerning the GRB–SN association is the attempt to check if all GRBs have SN counterparts. We have seen the difficulty of finding secure associa-
tions for distant GRBs \((z > 0.7)\). Although not all GRBs have revealed the signature of a SN, this does not indicate the absence of correlation because there are many reasons explaining why the SN detection in the afterglow may be difficult (see, e.g., Figure 6 of Woosley & Bloom 2006). However, there are some cases where technical problems cannot be invoked to explain the lack of SN light. This is clearly the case of GRB 060505 and GRB 060614 that are discussed in the following section.

8.6.3 Additional SN signatures and questions

The association of GRBs with active star-forming regions is also a good argument supporting the collapsar model. Hjorth et al. (2002) reported on the identification of the afterglow of GRB 980613. They presented late-time observations with HST (799 days after the burst) which made it possible to locate the GRB near a very compact blue object inside a region with star-forming knots and/or interacting galaxy fragments. Such examples reinforce the association of GRBs with very massive stars in star-forming regions.

Other evidence comes from observations of the host galaxies. They show very blue colors (Fruchter et al. 1999, Le Floc’h et al. 2003) and strong emission lines (OII, OIII) indicative of high star-formation rates (Vreeswijk et al. 2001) as expected in models involving the collapse of massive stars. Strong support for the collapsar model was also provided by a study by Bloom et al. (2002) concerning the locations of 20 long-duration GRBs within their host galaxies. A strong statistical connection of GRB locations with the UV light of their hosts was found, again providing observational evidence for the connection of GRBs with star-forming regions. In addition, the observed offset distribution was compared with the expected offset distribution of delayed merging remnants. This comparison ruled out delayed merging remnants as the progenitors of long-duration GRBs. Of course for many observations it is not possible to confirm the presence of the SN due to the weakness of its optical emission (especially at high redshift) and sometimes to the presence of the emission by the host galaxy (especially at low redshift). In fact it is really crucial to identify SNe with spectroscopic observations, but this is usually impossible for GRBs beyond redshift \(z = 1\). We have seen, however, that different types of signatures for the presence of a SN can be established; if they can be used together they certainly constitute strong support for the collapsar model.

Despite these very positive associations for long GRBs, various questions can be raised concerning the environment of GRBs, which arise from the time evolution of their afterglow light-curves. For instance if GRBs are coming from massive stars, the explosion has to take place in a medium enriched by constant mass-loss from the stellar winds. The signature of this wind-stratified medium would have to be present in the afterglow and revealed by a bright sub-millimeter emission at early times and by an increase of the cooling frequency with time (Chevalier & Li 2000). However, there are few unambiguous cases of GRBs occurring in such a medium and many afterglows are best fitted with a homogeneous surrounding medium. Another point concerns the observed densities: if GRB progenitors are situated in starburst regions, the expected densities should be higher than those usually inferred via the modeling of
the afterglow. In fact, the densities found in afterglow modeling are sometimes a factor of \( \sim 100 \) smaller than those expected for typical massive star winds (Chevalier, Li, & Fransson 2004). Of course, this is not to say that the progenitors are not associated with massive stars and that the collapsar is not a good model, but the issue remains and it has no completely satisfactory solution. It might be the indication that the surroundings of GRBs have been modified by stellar winds and/or ionizing radiation before the explosion. Scalo and Wheeler (2001) have emphasized that the winds and SNe occurring in a cluster of massive stars create super-bubbles whose density can be as low as \( 10^{-3} \text{ cm}^{-3} \). But they also noted that various circumstances may yield circum-burst medium densities which span a few orders of magnitude.

Also, one factor contributing to low density might be high-velocity winds found in some WO Wolf–Rayet stars; we have seen that this class of WR stars has been proposed as a GRB progenitor. The low metallicity of the environment might also lead to low densities. These points, which have been analyzed by Chevalier, Li, and Fransson (2004; see also Section 6.10) demonstrate the complexity of the problem.

These difficulties and the remaining uncertainties are illustrated by the discovery of two nearby long-duration bursts, which have challenged our previous views on the GRB–SN association: GRB 060505 at \( z = 0.089 \) and GRB 060614 at \( z = 0.125 \) (see Section 4.10.4). Fynbo et al. (2006) and Watson et al. (2007) reported that no supernova emission accompanied these two long-duration GRBs down to limits hundred of times fainter than the archetypal SN 1998bw (see Figure 4.24), and in fact fainter than any type Ic SN ever observed (Richardson, Branch, & Baron 2006). These two bursts originated in actively star-forming galaxies and, moreover, the possibility of significant dust obscuration is excluded. The absence of SN to such deep limits disagrees with previously known nearby long GRBs with optical afterglows, which were associated with SNe (but see the case of XRF 040701 at \( z = 0.20 \) which had no associated SN; Soderberg et al. 2005; and Section 4.5.2). These two events were found among six long GRBs or XRFs detected at very low redshift \( (z < 0.2) \). So the fraction of SN-less GRBs might be substantial. This naturally led Fynbo et al. (2006) to raise the question of a possible new type of massive stellar death. But before going to such an extreme solution it should be recalled that the early GRB model proposed by Woosley (1993) invoked ‘failed supernovae’, and this model was then adjusted when GRB-associated supernovae were discovered. The possibility of weak SNe had also been predicted and is perhaps demonstrated through these two GRBs (Fryer et al. 2007). Since the luminosity of a SN is roughly proportional to the total amount of \( ^{56}\text{Ni} \) produced in the explosion it would mean that some GRBs produce very little \( ^{56}\text{Ni} \) \( (<0.007 \ M_\odot) \). This would be the case for systems whose black hole forms via fall-back. In fact, in a collapsing star the black hole can be formed in two ways: direct collapse of the stellar core into a black hole or delayed collapse caused by fall-back in a weak SN explosion (Fryer, Young, & Hungerford 2006). Fall-back takes place when part of the material ejected after the SN explosion can decelerate and falls back onto the proto-compact object: a black hole or a neutron star. Fall-back is stronger when the explosion is weak, and it may be that this mechanism works only for weak SNe (Fryer et al. 1999). Other explanations for this absence of SNe have been discussed in Section 4.10.4.
To conclude on the SN–GRB connection, clear overviews on this topic, including observations, history, and theory, have been given by Soderberg (2006) and by Woosley and Bloom (2006). Soderberg (2006) compares GRB-associated SNe with local SNe of type Ibc. She selected three criteria for this comparison. The first criterion is the distribution of peak optical magnitudes versus $^{56}$Ni mass estimates. Figure 8.7 clearly shows the correlation and illustrates how the distributions of local and GRB-associated SNe overlap. No exceptional behavior of GRB-associated SNe is observable. Soderberg concluded that optical observations cannot be used to distinguish the class of GRB–SNe from the local SNe Ibc (see also Woosley & Bloom 2006). The second criterion is the radio luminosity at early time (a few days to weeks), which provides a measurement of the energy coupled to the on-axis relativistic ejecta. Here a distinction between the two classes of SNe appears. Radio-bright events like SN 1998bw are rare; less than 2% of the local population, and GRB–SN explosions are a factor of $\sim 10^3$ more radio-luminous than typical SNe Ibc. It can be deduced that central engines powering local SNe Ibc are much weaker than those powering GRB–SN explosions. The third criterion studied by Soderberg is the radio luminosity at late time (several years) which constrains the emission from all the relativistic material ejected by the supernova (on-axis and off-axis). Soderberg showed that none of the 68 Type Ibc supernovae in their sample had strong late-time radio emission that could be attributed to an off-axis GRB jet. She concluded that the fraction of SNe Ibc hosting GRB jets is lower than 10% (the majority of local SNe cannot have produced a GRB, even off-axis). Bissaldi et al. (2007) using different hypotheses concerning the progenitors of SNe Ibc, reproduced the local rate of SNe Ibc in galaxies and predicted a fraction of cosmic GRBs varying from 0.1% to 1% only. This is significantly smaller than the fraction calculated by Soderberg. They also predicted more GRBs at higher redshift. This large difference between local SNe Ibc and GRB–SNe might be due to special characteristics needed to produce GRBs, such as fast stellar rotation, asymmetric explosions and/or low metallicity. Soderberg (2006) emphasized the following point: SNe Ibc with broad optical absorption lines are not more luminous than other events and they do not imply the presence of relativistic ejecta. Finally, she concluded that, while most GRB explosions have a SN component (but see our previous reservations), only a small fraction of SNe Ibc, which remains ill-defined, produces the energetic relativistic ejecta which is the signature of GRBs.

8.7 NEUTRON STAR–NEUTRON STAR MERGERS

This is the second class of progenitors. As in the collapsar model a spinning black hole is formed. The debris from the tidally disrupted compact object forms a temporary accretion disk or torus, which ultimately falls into the black hole releasing a fraction of the gravitational energy of the matter. Therefore a black hole surrounded by a temporary torus of matter is the common endpoint of collapsars and mergers of compact objects. Mergers of compact objects can be very diverse: neutron star + neutron star (NS–NS), black hole + neutron star (BH–NS), black hole + white
dwarf (BH–WD), or black hole + helium star (BH–He). The class of coalescing compact binaries was introduced early on as a serious candidate for GRB progenitors (Paczynski 1986, Goodman 1986, Goodman, Dar, & Nussinov 1987, Eichler et al. 1989). At that time, the first binary pulsar PSR 1913+16 had already been discovered by Hulse and Taylor (1975) and the final destiny of binary systems of two neutron stars with very small orbital separation was predicted. The decrease of the separation between the two neutron stars due to gravitational radiation inevitably leads to the coalescence of the two neutron stars (Taylor 1994).

When it became evident that GRBs could be at cosmological distances, the rate of NS–NS coalescence was quickly estimated (see, for instance, Phinney 1991, Narayan, Paczyński, & Piran 1992, van den Heuvel & Lorimer 1996), providing estimates compatible with estimates of the GRB rate (Fenimore et al. 1993).

While NS–NS mergers are discussed most often, other types of compact binaries are also considered:

- BH–NS binaries, which can be formed at a rate comparable to and perhaps higher than that of NS–NS binaries (e.g. Bethe & Brown 1998).
- Other possible mergers, including BH–He binaries (Fryer & Woosley 1998) and BH–WD binaries (Fryer et al. 1999).

The formation scenarios of close binary systems of two compact objects have been reviewed by Fryer et al. (1999) and are reported in Figures 8.8, 8.9, and 8.10 (we have selected six scenarios proposed by Fryer et al. (1999) that illustrate the formation of NS–NS binaries (or DNS, for double neutron stars), BH–NS binaries, and BH–WD binaries).

We will see that the coalescence scenario is well adapted to explain the class of short-duration GRBs (hereafter SHBs, Dezalay et al. 1992, Kouveliotou et al. 1993) whose afterglows were reported for the first time by HETE-2 and mainly by Swift, but still for a small number of bursts (see Sections 4.4 and 4.11). We nevertheless note that Zhang, Woosley, and MacFadyen (2003) proposed a way to explain short GRBs in the collapsar model.3

In the following sections we consider mainly mergers of two compact stars, first mergers of two neutron stars then mergers of a black hole and a neutron star.

### 8.7.1 Mergers of two neutron stars

Of course, the merger of two neutron stars represents a huge reservoir of gravitational binding energy with the great advantage of a baryon clean region, which will appear along the rotation axis of the binary as a consequence of the huge centrifugal forces.

3 They suggested that a small amount of material could be pushed by the jet in the collapsar and accelerated to relativistic speeds when the jet breaks out. The interaction of this material with the stellar wind of the progenitor could produce a short transient emission (Woosley, Eastman, & Schmidt 1999). Zhang, Woosley, and MacFadyen (2003) speculated that this mechanism could be at the origin of hard precursors to long GRBs and, in some cases, could explain short, hard GRBs by themselves.
Figure 8.8. Two scenarios considered by Fryer, Woosley, and Hartmann (1999) for double (DNS) mergers with the final production of a GRB. MS is a main sequence star, $M_p$ and $M_s$ are the primary and secondary masses respectively. $M_{SN} \sim 10 M_{\odot}$ is the critical mass above which massive stars collapse to form supernovae, $M_{BH} \sim 25 M_{\odot}$ is the critical mass above which massive stars form a black hole.
Figure 8.9. Two scenarios considered by Fryer, Woosley, and Hartmann (1999) for BH–NS mergers with the final production of a GRB. MS is a main sequence star, $M_p$ and $M_s$ are the primary and secondary masses respectively, $M_{SN} \sim 10 \ M_\odot$ is the critical mass above which massive stars collapse to form supernovae, $M_{BH} \sim 25 \ M_\odot$ is the critical mass above which massive stars form a black hole.
Figure 8.10. Two scenarios considered by Fryer, Woosley, and Hartmann (1999) for BH–WD mergers with the final production of a GRB. MS is a main sequence star, $M_p$ and $M_s$ are the primary and secondary masses respectively, $M_{SN} \sim 10 M_\odot$ is the critical mass above which massive stars collapse to form supernovae, $M_{BH} \sim 25 M_\odot$ is the critical mass above which massive stars form a black hole.
A sudden energy release in this clean region can be quickly and continuously transformed into a radiation-dominated plasma as was very early noted by Cavallo and Rees (1978), Goodman (1986), and Paczyński (1986). The cleanliness of the fireball is critical because the baryon load must be lower than $\sim 5 \times 10^{-4} M_\odot$ for a fireball of $\sim 10^{53} \text{ erg}$ to reach Lorentz factors $\Gamma \sim 100$ or larger (see Section 5.1). We will see that the cleanliness of the fireball is a real issue, even for mergers of two compact objects.

The $e^+/e^-\gamma$ fireball arises from the enormous compressional heating and dissipation associated with the accretion; $\nu\bar{\nu}$ annihilation or magnetic fields in excess of $10^{14} \text{ G}$ can provide the driving forces leading to relativistic expansion. These two modes of energy extraction are very often considered, particularly by Rosswog, Ramirez-Ruiz, and Davies (2003). In NS–NS mergers a rapidly spinning black hole is formed, orbited by a neutron-rich high-density torus. Lee, Ramirez-Ruiz, and Page (2005) have given the physical conditions in the torus. The disk typically contains a few tenths of a solar mass; it is small, with the bulk of the mass being contained within 200 km of the black hole, which has a typical mass of 3 to 5 $M_\odot$. The disk is very dense: $10^9 < \rho < 10^{12} \text{ g cm}^{-3}$ with high internal energy ($10^{10} \text{ K} < T < 10^{11} \text{ K}$) (Rosswog et al. 1999, Lee & Ramirez-Ruiz 2002). The temperature is so high that nuclei are photo-disintegrated, leading to a mixture of $\alpha$ particles, free neutrons and protons, electrons and positrons. In this system the two main reservoirs of energy are the binding energy of the orbiting debris and the spin energy of the black hole. The extractable energy could amount to several times $10^{53}(M_\ast/M_\odot) \text{ erg}$ where $M_\ast$ is the mass of the torus. Rosswog, Ramirez-Ruiz, and Davies (2003) gave the amount of thermal and kinetic energy for the central object and for the debris: typical values for the central object are $2 \times 10^{52} \text{ erg}$ and $8 \times 10^{52} \text{ erg}$ in thermal and kinetic energy respectively, and for the debris the kinetic energy dominates by almost an order of magnitude, reaching $\sim 10^{52} \text{ erg}$ (vs $10^{51} \text{ erg}$ for the thermal energy).

In the following discussion of the coalescence of two neutron stars, we mainly consider two levels of analysis corresponding to two generations of models. The last part is a discussion about the galactic sites of these possible SHB progenitors.

(a) The first level is mainly centered on the work of Rosswog and collaborators and on the work of Ruffert and Janka. The framework of the work of Rosswog et al. is given in Rosswog and Davies (2002), Rosswog and Liebendörfer (2003), and Rosswog, Ramirez-Ruiz, and Davies (2003). They have presented 3D, high-resolution calculations of the last spiral-in stages and final coalescence of NS–NS binaries; the evolution of the system is followed for approximately 15 ms. They used Newtonian gravity, adding the back-reaction forces resulting from the emission of gravitational waves, and a new equation of state (EOS) based on the tables of Shen et al. (1998), which is stiffer than the Lattimer–Swesty EOS (1991). The work of Ruffert, Janka and collaborators is described in the following papers: Ruffert, Janka, and Schaefer (1996), Ruffert et al. (1997), and Ruffert and Janka (1999, 2001). Their simulations are also basically Newtonian; they take into account gravitational wave emission and the back-reaction on the
hydrodynamical flow, and they use the equation of state of Lattimer–Swesty in tabular form.

(b) The second level involves more detailed and more recent simulations considering a General Relativity approach, with time-dependent modeling extending over the typical duration of a SHB (short, hard GRB). This analysis will be completed by some recent results on simulations of binary neutron star mergers taking into account for the first time the magnetic field of the two stars and its enormous amplification which can occur within the first millisecond after the merger (Price & Rosswog 2006). This second level not only introduces a more sophisticated approach of a very difficult problem from the point of view of the hydrodynamics but also of the equation of state of the nuclear matter which characterizes the two stars.

8.7.2 NS–NS coalescence: the first generation of models

This analysis is largely based on the paper of Rosswog, Ramirez-Ruiz, and Davies (2003). These authors consider two options to transform the available energy into the outflowing relativistic plasma. A sketch of the two options is illustrated in Figure 1 of their paper, reproduced here in Figure 8.11. In Figure 8.11(a), neutrinos ($\nu$) give rise to a relativistic pair-dominated wind in the region of low baryon density. In Figure 8.11(b) strong magnetic fields anchored in the dense matter can convert the gravitational binding energy into Poynting-dominated outflow. The first mechanism involves neutrino and antineutrino pairs and the second, magnetic energy. As we have seen, the torus is hot and very dense so that only neutrinos can extract the large thermal energy which is present in the torus. It has to be noted that the total $\nu$ luminosities from the merger event calculated by Rosswog and Liebendörfer (2003) are much lower than in previous investigations, but we will return to this question. The neutrino emission is focused along the rotation axis, as the pole regions are surrounded by the high-density walls of the thick disk; the steep density gradient in the radial direction prevents lateral expansion; $\nu\bar{\nu}$ annihilation can accelerate a relativistic outflow in a region which is particularly baryon-clean. As noted by Rosswog, Ramirez-Ruiz, and Davies (2003) the exact densities in these areas are crucial, but very difficult to resolve numerically (Figure 8.12). The annihilation of $\nu\bar{\nu}$ pairs into $e^+/e^-$ pairs can tap the thermal energy produced by viscous dissipation in the torus. This mechanism, which can power a GRB, has been calculated by several groups (Mochkovitch et al. 1993, Jaroszynski 1996, Rosswog, Ramirez-Ruiz, & Davies 2003 and references therein). As we have said the disk geometry allows channeling the relativistic outflow into a pair of anti-parallel jets. This mechanism is similar to that considered by MacFadyen and Woosley (1999) for the collapsar scenario but the big advantage here is that the jet does not have to pierce the stellar envelope. The calculations of the ratio $\eta$ of the energy deposited via $\nu\bar{\nu} \rightarrow e^+, e^-$ to the rest mass $mc^2$ (which is a measurement of the maximum attainable Lorentz factor; see Section 5.1) are given in Figure 8.13 (Figures 4 to 6 of Rosswog, Ramirez-Ruiz, & Davies 2003). Of course due to the centrifugal evacuation of the funnels, the largest Lorentz factors are found along the rotation axis of the binary, giving the pair of
Figure 8.11. Illustration of the mechanisms which could produce a GRB from a NS merger (from Rosswog, Ramirez-Ruiz, & Davies 2003). (a) Neutrinos and anti-neutrinos from the hot remnant annihilate and produce a relativistic outflow—the fireball. The huge neutrino flux (i) drives an energetic baryonic wind (ii) that collimates the jet. (b) Strong magnetic fields anchored in the dense matter can convert the binding and/or spin energy into a Poynting outflow; dynamo processes are believed to operate in accretion disks, and the fields which are generated would have a typical length-scale of the order of the disk thickness. Open field lines that connect the disk to the outflow can drive a hydro-magnetic wind (i). This is a way to tap the binding energy of the debris torus (ii). At some stage after the merger a rapidly spinning BH is assumed to be formed whose energy can in principle be extracted through MHD coupling to the rotation of the hole (iii, iv). Magnetic fields which thread the BH are able to extract its rotational energy via the Blandford–Znajek process.

relativistic jets. It can also be seen in these figures that a severe decrease of $\eta$ appears at large angles from the rotation axis.

The typical energy deposited by $\nu\bar{\nu}$ annihilations is $10^{48}$ to $10^{49}$ erg, lower than the values given by Janka et al. (1999) and Lee and Ramirez-Ruiz (2002). These authors reported energies deposited by $\nu\bar{\nu}$ annihilations close to $10^{50}$ erg ($\nu\bar{\nu}$ annihilate with an efficiency around 1% and $\sim 10^{52}$ erg are released from the system in neutrinos, Janka et al. 1999). We will see that these low values were also obtained by
Figure 8.12. (a) Log(density—g cm\(^{-3}\)) along the rotation axis (z-axis) of the merged remnant of two 1.4 \(M_\odot\) neutron stars in corotation (from Rosswog & Liebendörfer 2003).

(b) Log(energy deposition—erg s\(^{-1}\) cm\(^{-3}\)) via \(\nu\bar{\nu}\) annihilation for the same simulation. The relativistic outflow is channeled along the z-axis where the baryon loading can be avoided. The only escape route is along the initial binary rotation axis. The steep density gradient in the radial direction prevents the lateral expansion of the outflow.

Ruffert et al. (1997), Janka & Ruffert (1996), and Ruffert and Janka (1999, 2001). The energies calculated by Rosswog, Ramirez-Ruiz, and Davies (2003) are obtained following the evolution of the merger during 0.2 s, comparable to the duration of a short burst. These authors have evaluated the energy deposition rate by \(\nu\bar{\nu}\) annihilation during the dynamical phase of the merging of binary neutron stars. With an annihilation efficiency of \(\sim 10^{-3}\) at the time of maximum neutrino emission, some 6–8 ms after the stars have merged, they found that little energy is deposited by \(\nu\bar{\nu}\) annihilation. In fact the integrated energy deposition during the duration of the simulation, \(\sim 10\) ms, is less than \(4 \times 10^{48}\) erg. So during this initial phase of merging, i.e. during the final stages of the inspiral of the two neutron stars and during the first few milliseconds after the merger, \(\nu\bar{\nu}\) annihilation is unable to produce enough energy to power a SHB at cosmological distances. But this phase of merging and the evolution that follows immediately afterwards do not take into account the possible role of a disk around the forming black hole. During the last phase, immediately following the merging, neutrinos can be emitted on a much longer timescale. Ruffert et al. (1997) considered a simple model for the post-merging emission from a disk (0.1 \(M_\odot\)) where viscous effects are considered as having a crucial influence on the disk evolution and \(\nu\) emission. They determined a value of the shear viscosity in the disk, called \(\eta^*\), where \(\nu\bar{\nu}\) annihilation yields the largest energy deposit. They calculated the total neutrino luminosity and the total energy \(E_\nu\) during the accretion timescale of the disk, considering this optimum value \(\eta^*\) and taking into account a \(\nu\bar{\nu}\) annihilation efficiency \(\varepsilon_{\nu\bar{\nu}}\) of a few \(10^{-3}\) (their relation 10). The total energy \(E_{\nu\bar{\nu}} = \varepsilon_{\nu\bar{\nu}} E_\nu\) deposited in an \(e^+/e^-\) pair photon fireball by the annihilation of \(\nu\) and anti-\(\nu\) radiated in the disk is \(E_{\nu\bar{\nu}} \propto 10^{49}–10^{50}\) erg (see also Jaroszynski (1996) for similar values). For these
Figure 8.13. Ratio of the energy deposited via neutrino annihilation ($\nu\bar{\nu} \rightarrow e^+e^-$) to baryon rest mass energy for the region above the poles of the merged remnant. This is a measure of the maximum attainable Lorentz factor. Due to the centrifugal evacuation of the funnels, the largest Lorentz factors are obtained along the binary rotation axis. At large angles from the rotation axis an increasing degree of entrainment leads to a drastic decrease of the Lorentz factor. Starting from the top three models are considered (from Rosswog, Ramirez-Ruiz, & Davies 2003). (a) Merger of two 1.4 $M_\odot$ non-rotating neutron stars 18.3 ms after the merger, with neutrino cooling. (b) Merger of two 1.4 $M_\odot$ corotating neutron stars 10.7 ms after the merger, with neutrino cooling. (c) Merger of two 2.0 $M_\odot$ non-rotating neutron stars 18.3 ms after the merger, with neutrino cooling. See also Color section.
calculations a disk of \(0.1 \, M_\odot\) is considered around a \(3 \, M_\odot\) black hole. These values are significantly larger than the \(10^{48}\) erg calculated for the dynamical phase of the merging which lasts a very short time, \(\sim 1\) ms. In the post-merging evolution, the disk emits \(\nu\) with similar luminosities but on a much longer timescale: a few hundred milliseconds. These two different timescales explain the significant difference in the energy available for the fireball. This mechanism can explain weak bursts with energy \(\sim 10^{50}\) erg and durations shorter than or around 1 s (typical of SHBs), but the annihilation energy \(E_{\nu\bar{\nu}}\) remains too low (a factor of 10) to explain powerful, longer cosmological GRBs. For instance, the two short bursts GRB 050724 and GRB 051221 have beaming corrected energies which are respectively \(4 \times 10^{49}\) erg and \(>1 \times 10^{50}\) erg (see Section 4.11) and this last value is the largest of the four SHBs reported with an afterglow detection, so their energy can be explained by \(\nu\bar{\nu}\) annihilation in the disk of a binary merger.

Even though the energy is available, a major concern raised by Ruffert et al. (1997) comes from the baryonic wind, which is blown off the surface of the merger and of the accretion torus by neutrino heating. For Ruffert and Janka (1998) the baryon loading of the \(e^+ / e^-\) photon fireball is at least five orders of magnitude too high, preventing the formation of a highly relativistic expanding fireball, which could lead to a GRB. On the other hand a positive point (see their equation 26) is the dependence of \(E_{\nu\bar{\nu}}\) on the mass of the black hole; \(\nu\bar{\nu}\) annihilation increases steeply with black-hole mass: \(E_{\nu\bar{\nu}} \propto R_\bullet^{13/8} \propto M^{13/8}\). Thus higher black-hole masses can lead to significantly higher energy available for the fireball. Ruffert and Janka (1999) presented a new and complete study of the formation and properties of the accretion torus around a stellar mass black hole, the remnant of a neutron star merger. They evaluated different neutron star merger models for the energy deposition by \(\nu\bar{\nu}\) annihilation around the accretion torus. They found that the torus has a mass between several \(10^{-2} \, M_\odot\) to a few \(10^{-1} \, M_\odot\) and maximum density around \(10^{12} \, g \, cm^{-3}\), the maximum temperature being \(\sim 10\) MeV. Neutrino–antineutrino annihilations deposit energy in the vicinity of the torus at a rate \((3–5) \times 10^{50}\) erg s\(^{-1}\). This corresponds to an efficiency of \(0.4–0.5\%\) for the conversion of \(\nu\) energy into \(e^+ / e^-\) pairs. Near the rotation axis they found that \(10–30\%\) of this energy (up to \(10^{49}\) erg) is released within an estimated emission period of 0.02 to 0.1 s in a region with a low baryonic mass, typically \(10^{-5} \, M_\odot\). This energy is sufficient to power SHBs with moderate beaming. If the beaming factor \(f_\Omega\) is \(10^{-2}\) to \(10^{-1}\) (opening half-angles of \(10^\circ\) to several tens of degrees) \(\gamma\)-ray burst energies \((E_\gamma \sim E_{\nu\bar{\nu}} / f_\Omega)\) in the range \(10^{50}–10^{51}\) erg can be produced by neutron star mergers. This confirms the results of previous studies and conclusions of Ruffert et al. (1997; see also Popham, Woosley, and Fryer (1999) who found \(5 \times 10^{49}\) erg for the total energy available for a burst). All the reservations and remarks concerning the limitations of these simulations and the possible various improvements are discussed in their conclusion where they indicate that these analyses still have to be considered with caution.

Ruffert and Janka (2001) presented a compilation of results from their most advanced neutron star merger simulations, confirming that the improved models arrive at the conclusion we have just presented: GRBs cannot be powered by neutrino emission during the dynamical phase of the merging of the two neutron stars. With an
extended computational grid the neutrino luminosities are found to be a factor of 2–4 larger (a few $10^{53}$ erg s$^{-1}$) than in previous models (Ruffert et al. 1997). It follows that the energy deposition rates by $\nu\bar{\nu}$ annihilation in the dilute outer layers of the post-merging object are up to a factor of 30 larger, but the authors consider that since this energy is transferred to a region with large baryonic density, electrons and positrons are captured on nucleons with $\nu$ emission. So this energy is inefficient in powering a GRB and should drive a baryonic wind rather than a relativistic outflow of a baryon-poor pair photon plasma. Janka and Ruffert (1996) previously indicated that 75–80% of the annihilation energy is deposited in or near the disk and serves to drive a baryonic wind. This decreases considerably the efficiency of $\nu\bar{\nu}$ annihilation to produce $e^+/e^-$ pairs in a baryon-free funnel and therefore the energy available to produce energetic $\gamma$-rays. For these authors, considering either the dynamical phase of the merging of two neutron stars or the post-merging phase, GRBs cannot be powered via $\nu\bar{\nu}$ annihilation except if they are weak and/or if the focusing of the fireball is strong. This could be the case because the barrier of the dense baryonic gas of the torus could lead to beaming of the $e^+, e^-, \gamma$ plasma. Roesswog and Ramirez-Ruiz (2003) considered the collimation of the $\nu\bar{\nu}$ annihilation fireball by the $\nu$-driven wind; the existence of such a baryonic wind is a natural consequence of the huge gravitational energy released in the form of $\nu$. They show that this wind, driven from the disk, could have enough pressure or inertia to provide the collimation of the fireball. Thus SHBs produced by this mechanism would not be isotropic but beamed within a typical solid angle of $\sim 0.1$ sr (see also Aloy, Janka, and Müller (2005) and the discussion of their work below). These simulations are done for a typical duration of 15 ms. Moderate beaming of ten to a few tens of degrees (the opening half-angle of the beam) leads to GRB isotropic energies of the order of $10^{50}$ to $10^{51}$ erg. Here $E_\gamma \sim E_{\nu\bar{\nu}}/f_\Omega$ with $f_\Omega$ between $10^{-2}$ and $10^{-1}$. Under these conditions short GRBs with luminosities reaching $10^{51}$ erg s$^{-1}$ for typical burst durations of 0.1 to 1 s can be produced by neutron star mergers at least during the post-merging emission from the disk.

Later, Setiawan, Ruffert, and Janka (2004, 2006) considered the evolution of the BH–torus system on a longer timescale (40 ms), which is comparable with the duration of the shorter SHBs. They carried out 3D hydrodynamic simulations using a pseudo-Newtonian potential to approximate the effects of general relativity and they used the EOS of Lattimer and Swesty; they took into account viscous dissipation but ignored magnetic fields. They investigated the time-dependent evolution and properties of the accretion torus around rotating and non-rotating stellar mass black holes for 40 ms. The black hole and its torus are considered to be the remnants of a binary neutron star or of a black hole–neutron star merger; the black-hole mass is $4 M_\odot$ and the mass of the accretion disk is 0.01 $M_\odot$ to 0.2 $M_\odot$. They found that the time-dependent efficiency for converting gravitational energy into neutrinos, expressed by the ratio of the $\nu$ luminosity to the rest mass energy of the accreted material, can reach values up to $\sim 10\%$. The efficiency of converting $\nu$ energy into a pair–photon fireball by $\nu$ annihilation peaks at several percent. For a sufficiently massive torus ($>0.1 M_\odot$), high disk viscosity ($\alpha \sim 0.1$) and BH spin parameter ($a = 0.6$), the neutrino annihilation can thus deposit a total energy of some $10^{49}$ erg above the poles
of the BH (Setiawan, Ruffert, & Janka 2004). These authors also found that the \( \nu \) luminosity and total energy release of the torus increase steeply with higher viscosity, larger torus mass, and larger black-hole spin. Their conclusion is that for typical post-merger rotation rates and viscosity with \( \alpha \sim 0.1 \), and for the torus masses investigated, sufficient energy can be released to account for SHBs, if beaming fractions of about 1% are considered (see, however, Lee, Ramirez-Ruiz, and Page (2004) for a comment on the accretion timescale at this high viscosity, which is too short). Under these conditions Setiawan, Ruffert, and Janka (2006) concluded that there is no need to invoke additional power from mechanisms based on the presence of extremely strong magnetic fields.

Nevertheless, the role of magnetic processes had been studied very early, because they represent an alternate way to tap the energy contained in the torus. In the following discussion we will refer to the paper of Rosswog, Ramirez-Ruiz, and Davies (2003).

Even before a black hole forms, the merging system might lead to winding up the fields with energy dissipation before the merger. This mechanism can extract the rotational energy available in the torus. When the black hole is formed it is rapidly rotating (~millisecond period) and very massive, and it represents a larger energy reservoir than the torus. Magnetic configurations that can power GRBs require fields of a few \( 10^{15} \) G. These huge fields can be quickly reached. As shown by Kluźniak and Ruderman (1998) a simple winding up of field lines by differential rotation is able to amplify a \( 10^{12} \) G field to \( 10^{15} \) G very rapidly. This corresponds to the DROCO mechanism analyzed by Kluźniak and Ruderman (1998) (DROCO stands for differentially rotating collapsed objects as central engines for GRBs). The onset of the neutron star collapse is expected to initiate differential rotation, i.e. the birth of a short-lived DROCO and the occurrence of a GRB if the DROCO phase preceded the collapse into a BH. The accretion-induced collapse of a NS into a BH may also give rise to a DROCO phase. The DROCO may also be a torus rotating about a spinning BH. The DROCO will wind up initial seed magnetic fields by differential rotation until the magnetic pressure \( B^2/8\pi \) becomes comparable to the pressure of the matter. At this point magnetic toroids will become buoyant, float up and break through the surface where they will create an ultra-relativistic blast. Figure 8.14 illustrates simply the magnetic development in a DROCO. Rosswog et al. (2003) considered other mechanisms: an effective \( \alpha-\Omega \) dynamo (Thompson & Duncan 1993) or a Magneto-Rotational Instability (MRI, Balbus & Hawley 1998). These mechanisms can amplify a seed field exponentially on a much shorter scale than the previous one (DROCO). If such huge magnetic fields thread the black hole, its rotational energy represents another source of energy via the Blandford–Znajek (BZ) mechanism (Blandford & Znajek 1977). For a maximally rotating black hole this energy is \( 0.29 M_{\text{BH}} c^2 \) erg multiplied by an efficiency factor \( \varepsilon \). With a value of 20% for \( \varepsilon \), a jet powered by the BZ mechanism might produce the equivalent of an isotropic energy of \( \sim 10^{53} \) erg. As the entrained baryonic mass in these coalescence events is very low along the rotation axis, certainly below \( 10^{-4} M_\odot \), a highly relativistic fireball could develop and produce a short GRB, even without the need for beaming (Rosswog & Ramirez-Ruiz 2002). The conclusion is that the merger remnant with its large seed magnetic fields and its
turbulent fluid motion can generate what Rosswog, Ramirez-Ruiz, and Davies (2003) call a plethora of electromagnetic activity: coronal arches with reconnection, magnetic field amplification due to differential rotation and BZ processes when magnetic connection exists between the disk and the hole (Figure 8.11b). Therefore, their conclusions confirm that the coalescence of two neutron stars can power short GRBs through the $\nu\bar{\nu}$ process and perhaps better through the huge available magnetic energy, extracting the available energies at rates in excess of $10^{52}$ erg s$^{-1}$ for possible duration as long as 1 s. The requirements on the equivalent isotropic energies of short GRBs could be fulfilled even without beaming.

Before analyzing more recent and complete work, it is interesting to note that during this period a lot of other studies focused on $\nu\bar{\nu}$ annihilation in hyper-accreting black-hole systems with the steady-state structure and composition of post-merger $\nu$-cooled accretion disks (Popham, Woosley, & Fryer 1999, Narayan, Piran, & Kumar 2001, Kohri & Mineshige 2002, Di Matteo, Perna, & Narayan 2002, Beloborodov 2003). These authors introduce an increasing amount of microphysical details in the numerical modeling, concerning the equation of state of the neutron stars and the neutrino processes (see also Ruffert & Janka 2001, Rosswog & Liebendörfer 2003, Rosswog, Ramirez-Ruiz, & Davies 2003). These works, as noticed by Aloy, Janka, and Müller (2005), have given a theoretical understanding of the conditions which are present in the vicinity of the accreting black hole and
which determine the efficiency of energy losses by $\nu$-emission and the efficiency of energy conversion by neutrino annihilation. They complete the other works we have analyzed. For instance, Popham, Woosley, and Fryer (1999) and Narayan, Piran, and Kumar (2001) have shown that accretion onto a massive black hole (a few solar masses) at a rate of a fraction to a few solar masses per second proceeds via $\nu$ cooling in a neutrino-dominated accretion disk (NDAF). They found that, for such hyper-accreting systems, a total energy of $\sim 5 \times 10^{49}$ erg can be available for a burst, in agreement with the typical values which we have already reported. Taking into account opacities and neutrino transfer (see also Ruffert & Janka 1999), Di Matteo, Perna, and Narayan (2002) showed that for accretion rates larger than 1 $M_\odot$ s$^{-1}$ the optical depth of the accreting gas is enormous and neutrinos are sufficiently trapped so that energy cannot escape. The neutrinos are advected into the black hole and energy advection becomes the dominant cooling mechanism in the inner regions of the flow. Under these conditions $\nu\bar{\nu}$ annihilation in hyper-accreting black holes might be inefficient for liberating large amounts of energy. Both authors have noted the importance of the $\nu$ opacity at large accretion rates, $M > 1 M_\odot$ s$^{-1}$, and more generally the sensitivity to the accretion rate of the production of a pair fireball along the rotational axis by $\nu$ annihilation. Thus, Di Matteo, Perna, and Narayan (2002) stressed the importance of the extraction of rotational energy in the disk by magnetic processes as the most viable mechanism, the differential rotation of the disk amplifying pre-existing magnetic fields as we have already indicated. Kohri and Mineshige (2002) studied massive accretion disks dominated by $\nu$ cooling at huge accretion rates. The properties of such disks are studied based on the $\alpha$ viscosity model with particular attention to their thermal structure and to the possible impact of huge magnetic fields.

Let us now look at the most recent studies of NS–NS mergers.

### 8.7.3 NS–NS coalescence: the second generation of models

Understanding the evolution of BH–torus systems requires time-dependent modeling over durations longer than the typical 10 to a few tens of milliseconds considered in the simulations that we have already mentioned. This is necessary if one wants to approach the typical duration of short hard bursts, which is a few 100 ms. We will complete the previous analysis of NS–NS coalescence by reporting on a new generation of time-dependent studies which are very useful for answering or contributing to answer the following questions:

- What is the mass remaining in the accretion disk beyond the dynamic timescale?
- What is the mass accretion rate of the black hole vs. time?
- How do the properties of the accretion torus evolve with time: density, temperature, composition, $\nu$ luminosity?
- What is the influence on these properties of the torus mass, of the gas viscosity, and of the black-hole rotation?
- What are the $\nu$ luminosity and the efficiency of the $\nu\bar{\nu}$ annihilation for powering ultra-relativistic outflows and GRBs?

Some of these points have already been considered in the studies reported above but if one wants to go further it is essential to use time-dependent modeling and General Relativistic hydrodynamics. They have been studied recently and we will now report on some of the contributions of these simulations, which are essential for confirming whether NS–NS mergers are serious candidates for the central engines of SHBs.

**Time-dependent modeling**

Before the study of Lee, Ramirez-Ruiz, and Page (2004), the dynamics of the problem had never been followed for an amount of time comparable with the typical duration of SHBs: $\sim 0.2$ s. In the work of Setiawan, Ruffert, and Janka (2004), for instance, the calculations were stopped at $\sim 40$ ms due to computational limitations. Lee, Ramirez-Ruiz, and Page (2004) extended dynamical simulations of post-merger accretion disks around black holes to $\sim 1$ s. They used a realistic EOS and, for the first time, took into account the effects of $\nu$ opacities. The limitation on the cooling rate of the disk imposed by the high optical depths results in energy losses that are spread out over an extended period of time. The effect of neutrino opacity on the timescale of energy extraction may be essential for determining the duration of neutrino-driven short GRBs (Lee, Ramirez-Ruiz, & Page 2004). The viscosity, characterized by the coefficient of viscosity $\alpha$ (which allows a parameterization of the viscous stresses and energy dissipation rates; Shakura & Sunyaev 1973) has a major impact. For high viscosity ($\alpha \sim 0.1$) the accretion timescale is so short (40 ms) that most of the material is accreted onto the black hole before it has the time to radiate away its internal energy. On the other hand, at lower viscosities ($\alpha \sim 10^{-2}$ and $10^{-3}$, the two other cases considered by Lee, Ramirez-Ruiz, & Page (2004)), the accretion timescale is longer, 0.4 s and 1 s respectively. Hence the material is allowed to remain longer in the vicinity of the black hole before falling into it. This extended period can be long enough for the torus to radiate away essentially all of its internal energy through the emission of neutrinos, the $\nu$ opacity being essential to characterize at what rate the disk will be cooled. The total energy released in neutrinos is around $10^{52}$ erg, depending on the viscosity parameter $\alpha$. If this energy is converted into $e^+e^-$ pairs with an efficiency of 1% in the region along the rotational axis, around $10^{50}$ erg are available to produce a relativistic fireball. In order to attain an apparent isotropic energy of $10^{52}$ erg the fireball has to be collimated into a solid angle $\Delta\Omega \sim 10^{-2}$. To avoid this possible difficulty Lee, Ramirez-Ruiz, and Page (2004) also calculated the energy which can be extracted by MHD processes through the Blandford–Znajek mechanism and through a magnetically dominated jet assuming energy equipartition between the internal energy density $\rho c^2$ ($\rho$ is the matter density) and the magnetic energy density $B^2/8\pi$. The inferred magnetic field is around $10^{15} - 10^{16}$ G. They examined the possibilities of generating such strong magnetic fields. They estimated a total energy release by the BZ mechanism which is between $3 \times 10^{50}$ and $6 \times 10^{51}$ erg, for $\alpha$ varying from 0.1 to $10^{-3}$. This energy is liberated on a timescale which goes from
0.4 s for a high-viscosity disk to more than 1 s in the case with the lowest viscosity
($\alpha \sim 10^{-3}$) (Lee, Ramirez-Ruiz, & Page 2004). Figure 8.15 (from Lee 2005) gives the
$\nu$ luminosity and the BZ luminosities versus time for different values of the disk
viscosity ($\alpha$ parameter) and for different initial masses of the disk. So even if the $\nu$\bar{$\nu$}
processes require strong beaming, which seems possible, in any case the magnetic
processes make it possible to easily obtain the required energy for short GRBs, and
therefore mergers of two neutron stars seem to be realistic progenitors for this type of
GRB. A word of caution, noted by Lee, Ramirez-Ruiz and Page (2005) is that a small
mass fraction of baryons loading the outflow can severely limit the attainable Lorentz
factor; for instance, a Poynting flux of $10^{53}$ erg cannot accelerate an outflow to
$\Gamma > 100$ if it has to drag more than $10^{-4} M_\odot$ of baryons with it. Finally, Lee

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{figure8.15.png}
\caption{Upper panel: Neutrino luminosity $L_\nu$ for disk masses 0.3 and 0.06 $M_\odot$
(solid curves, from top to bottom) and for different viscosities characterized by the
parameter $\alpha$ varying from $10^{-1}$ to $10^{-3}$:
$\alpha = 0.1$ solid lines,
$\alpha = 0.01$ long dashed lines, $\alpha = 0.001$ short-dashed lines. The energy
is released over a timescale given by
$t_{\text{cool}} \sim E_{\text{int}}/L_\nu$ typically
0.1 s. $E_{\text{int}}$ is the internal energy reservoir of the
disk. Lower panel:
Luminosity $L$ when the outflow is powered by
the Blandford–Znajek mechanism for the same
values of $\alpha$ and disk masses. The decay
occurs on an accretion timescale $t_{\text{acc}}$ typically
$\sim t_{\text{burst}}$. (From Lee 2005.)}
\end{figure}
(2005) noted that these calculations were not performed using General Relativity and that explicit magnetic field effects were not taken into account in the hydrodynamics. He noted that there was still a long way to go from modeling the structure and evolution of the inner disk at the heart of the GRB central engine to the observed electromagnetic signal.

**General Relativity hydrodynamics**

Significant progress was obtained by taking into account fully general relativistic (GR) hydrodynamics and physically realistic nuclear equations of state (EOS). Aloy, Janka, and Müller (2005), for instance, developed the first general relativistic hydrodynamic models of the launch and evolution of relativistic jets and winds driven by thermal energy deposition ($\nu\bar{\nu}$ annihilation) in the vicinity of an accreting black hole–torus system. Through their simulations they try to answer critical questions connected with the production of SHBs, like the collimation mechanism of the polar outflow, the opening angle of the ultra-relativistic ejecta, the Lorentz factors which are involved, the duration of possible GRBs, and the isotropic equivalent energy. Aloy, Janka, and Müller (2005) showed that ultra-relativistic outflows with high terminal Lorentz factors ($\sim 1000$) can develop and that their interaction with the dense accretion torus leads to the collimation of the outflow into a sharp-edged core ($5^\circ–10^\circ$) embedded laterally in a wind with steeply declining Lorentz factor. The models proposed by the authors span a range of one hundred in isotropic equivalent energy $E_{\text{iso}}$, from $\sim 10^{49}$ to $10^{51}$ erg. This work confirms the conclusions that we have already presented: post-merger black hole–torus systems are viable engines of SHBs, and $\nu\bar{\nu}$ annihilation appears to be a sufficient energy source. The model can also explain the duration of all observed SHBs (Aloy, Janka, & Müller 2005). The collimation, $\theta_{\text{jet}} \sim 14^\circ$, found by Fox et al. (2005) for GRB 050709 seems to agree with these simulations, meaning that only 1% of all SHBs point to Earth. Of course, many more SHBs are needed before we can make definite statements about this topic (see also Section 8.10).

Janka et al. (2006) have used these relativistic hydrodynamics simulations to study the properties of off-axis SHBs. For these bursts the polar jet in binary mergers does not have to plough through many solar masses of overlying stellar matter as in the case of the collapsar model (where the black hole–torus system is at the center of a massive star). As the acceleration is not damped by swept-up matter, the jet very quickly reaches Lorentz factors of a few. The collimation of the baryon poor jets is provided by the much denser torus walls which surround the empty polar regions of the black hole. The collimated ultra-relativistic outflows possess flat core profiles, which are bounded at their lateral edges by very steep gradients. The rapid decrease of the isotropic equivalent energy as a function of $\theta$ implies that the probability of observing the lateral, lower Lorentz factor wings is expected to be significantly reduced (Janka et al. 2006). The observability of SHBs is thus strongly favored within a cone of opening semi-angle between 10$^\circ$ and 15$^\circ$ around the axis of the jet. These authors also found that the steep wings of the jet profiles might explain why SHBs are typically harder than long ones: higher Lorentz factors and steep jet edges may be the
characteristic features of GRB jets that originate from post-merger black hole–torus systems. Another point which is raised by Janka et al. (2006) concerns the X-ray flares detected in some SHBs (see Section 4.11). They might be explained by an extended period of reduced source activity, following the prompt emission, in which X-ray flares might be powered by a magnetic mechanism (Fan, Zhang, & Proga 2005). This work is still exploratory because observational data are too limited and the theoretical work has to be considered as first steps. As noted by Aloy, Janka, and Müller (2005) their relativistic hydrodynamic flow is triggered by the deposition of energy near the black hole–torus system, but this energy release is prescribed according to a defined functional behavior. To avoid this limitation the authors are working on two-dimensional relativistic simulations of the viscosity-driven evolution of the system with treatment of neutrino transport and $\nu\bar{\nu}$ annihilation (Aloy, Janka, & Müller 2005).

Shibata, Taniguchi, and Uryū (2005) have also used simulations in full general relativity and they also stressed the need to have a realistic EOS. One of their conclusions concerning GRBs is that in black-hole formation most of the mass is swallowed into the horizon and hence the disk mass is much smaller than 1% ($10^{-2}$–$10^{-3} M_\odot$). Although the disk is hot, with a temperature of 10–20 MeV, the thermal energy available for the neutrino emission is expected to be at most $\sim 10^{50}$ erg. In addition, the efficiency for converting the $\nu$ energy into electron–positron pairs by $\nu\bar{\nu}$ annihilation would be $< 10^{-4}$ (Ruffert et al. 1997). Thus it seems to be very difficult to generate cosmological GRBs with such mergers of binary neutron stars. But this result depends on the ratio of the masses of the mergers: the disk has much lower mass and thus insufficient energy to power a SHB if the merger masses are near equal: $q = M_2/M_1 \sim 1$ (where $M_1$ is the mass of the primary star; Shibata, Taniguchi, and Uryū 2005). Shibata and Taniguchi (2006) considered other initial conditions and especially the case with $q$ lower than 1 ($q \sim 0.7$). They recall the two scenarios which are identified to lead to the production of short, hard bursts (SHBs) through the NS–NS merger.

- In the first scenario, a hyper-massive neutron star (HMNS) is formed first. The angular momentum is quickly transported from the inner region to the outer envelope, which subsequently forms a hot torus, and later the HMNS eventually collapses to a black hole.
- In the second scenario a black hole is promptly formed, a torus is then formed from the less massive neutron star which is tidally elongated at the merger and subsequently constitutes a hot and geometrically thick torus.

These two scenarios are summarized in Figure 8.16 (their Figure 21).

**Scenario 1**

If the total binary mass is below some critical threshold $M_{\text{thr}}$ ($\sim 2.8–2.9 M_\odot$), a differentially rotating hyper-massive neutron star (HMNS) can be formed, initially stable against gravitational collapse (for more than 10 ms), in spite of a mass larger
Figure 8.16. Summary proposed by Shibata and Taniguchi (2006) giving the outcome after the merger. Acronyms HMNS, GW, B-field and J-transport respectively mean: hyper-massive NS, gravitational wave emission, magnetic field, and angular momentum transport. Small disk, massive disk and heavy disk refer to disk masses $M_d \ll 0.01 M_\odot$, $0.01 < M_d < 0.03 M_\odot$, and $M_d > 0.05 M_\odot$, respectively. $Q_M$ is the mass ratio of the system, the ratio of the rest-mass of the secondary to the primary star. $M_{\text{thr}}$ is the threshold mass for prompt formation of a BH, $M_{\text{thr}} \sim 2.8$–$2.9 M_\odot$; if $M < M_{\text{thr}}$ a hypermassive NS is formed at the outcome of the merger. Solid arrows denote the route found by the authors and dashed arrows on the right are the possible routes based on the results of Shibata et al. (2006). Hence two ways seem possible to produce short GRBs either through direct BH formation or through the formation of hyper-massive neutron stars (from Shibata & Taniguchi 2006).

than the maximum possible mass for uniformly rotating configurations. It is the differential rotation which allows exceeding the super-massive limit obtained for rigidly rotating stars for a given EOS (Baumgarte, Shapiro, & Shibata 2000). The angular momentum is transported from the inner region to the outer envelope which subsequently forms a hot disk (torus). In the meantime the HMNS may collapse to a black hole either by angular momentum dissipation due to gravitational radiation or by angular momentum transport probably due to magnetic effects (Baumgarte, Shapiro, & Shibata 2000). In this scenario the estimated mass of the torus is $\sim 0.01$–$0.03 M_\odot$ in the absence of strong magnetic effects and it can be $>0.05 M_\odot$ in its presence (Duez et al. 2006, Shibata et al. 2006). For instance Duez et al. (2006) tracked the evolution of a magnetized HMNS in full general relativity for the first time. Even with an initial weak seed magnetic field, this field can grow to influence significantly the dynamics of the system. The amplification of the magnetic field is due to magnetic winding and to the magneto-rotational instability (MRI) in the presence of differential rotation (Velikhov 1959, Balbus & Hawley 1998). Duez et al. (2006) found that secular angular momentum transport due to magnetic braking and to the magneto-rotational instability (MRI) results in the collapse of the HMNS into a rotating black hole. The black hole is surrounded by a hot massive torus undergoing quasi-stationary accretion and a collimated magnetic field. They suggest that the collapse of the HMNS is a possible candidate for the central engine of SHBs. Shibata et al. (2006) also explored this possibility using magneto-hydrodynamic simulations in full general relativity. They considered a magnetized HMNS undergoing a delayed
collapse to a rotating black hole with angular momentum transport via magnetic braking and magnetorotational instability (MRI). They found that a black hole is formed, surrounded by a massive hot torus with a collimated magnetic field. The torus accretes onto the black hole at a quasi-steady accretion rate \( \dot{M} \sim 10 M_\odot \text{ s}^{-1} \), and its lifetime is \( \sim 10 \text{ ms} \). With a temperature of \( 10^{12} \text{ K} \) the copious production of \( \nu \bar{\nu} \) and their subsequent annihilation leads to a total energy of \( \sim 10^{48} \text{ erg} \) in \( \gamma \)-rays, which may be sufficient to power an SHB if the emission is somewhat beamed, which is quite possible as we have seen (Aloy, Janka, & Müller 2005). The conclusions are the following: HMNS formation is possible in most merging NS–NS binaries; MHD simulations in full GR show that the HMNS undergoes a delayed collapse resulting in a hot, magnetized torus surrounding a rotating black hole, which presents a magnetic field collimated along the polar axis. Such merging binaries can produce SHBs, the burst being powered either by neutrino annihilation or by MHD effects.

**Scenario 2**

This scenario, also considered by Shibata and Taniguchi (2006) concerns higher-mass neutron stars or mergers with a binary mass ratio sufficiently far from unity (see Figure 8.16). These systems lead to the formation of a relatively massive torus around a black hole, which is formed promptly during the merger. The torus is formed from the less massive neutron star, which is tidally elongated at the merger and, after that, constitutes a hot, geometrically thick torus. The mass of the torus strongly depends on the value of \( q (= M_2/M_1 < 1) \), where \( M_1 \) is the mass of the primary star) and on the nuclear EOS of the neutron star. For mass ratios \( q < 0.7 \), tori with masses \( M_t \) from 0.01 to 0.1 \( M_\odot \) and mass accretion rates from \( \dot{M} \sim 0.1 \) to 1 \( M_\odot \text{ s}^{-1} \) can be obtained. With such values, the lifetime of the torus is \( t \sim 10–1000 \text{ ms} \), compatible with the durations of SHBs. Due to viscous dissipation in the torus, the kinetic energy is converted into thermal energy. Shibata and Taniguchi (2006) found that the total thermal energy produced during the accretion reaches \( 10^{51–52} \) erg. Only a fraction of this energy has to be transferred to the fireball to produce a SHB. Different scenarios have been proposed by Janka and Ruffert (1996), Ruffert et al. (1997), Popham, Woosley, and Fryer (1999), Narayan, Piran, and Kumar (2001), Ruffert and Janka (2001), Di Matteo, Perna, and Narayan (2002), Kohri and Mineshige (2002), Setiawan, Ruffert, and Janka (2004), and Lee, Ramirez-Ruiz and Page (2005). Due to its high temperature (typically \( 10^{11} \text{ K} \)) and density (\( \sim 10^{11} \text{ g/cm}^3 \)), the torus radiates strongly in thermal neutrinos. Numerical results (Shibata & Taniguchi 2006) suggest the formation of a hot, hyper-accreting torus optically thick to neutrinos. For \( \dot{M} \sim 0.1–1 M_\odot \text{ s}^{-1} \), the luminosity due to \( \nu \bar{\nu} \) annihilation is: \( L_{\nu\bar{\nu}} \sim 10^{49–50} \text{ erg s}^{-1} \) but for smaller \( \dot{M} (\dot{M} < 0.1 M_\odot \text{ s}^{-1}) \) the luminosity steeply decreases below \( 10^{49} \text{ erg s}^{-1} \). Aloy, Janka, and Müller (2005) simulated the propagation of jets powered by energy input along the rotation axis. They found that if the half-opening angle of the energy injection region is relatively small (<45°) and if the baryon density around the black hole is sufficiently low, jets with Lorentz factors in the hundreds can be produced with an energy input \( L_{\nu\bar{\nu}} \sim 10^{49–50} \text{ erg s}^{-1} \) lasting typically 100 ms. The conversion rate from the pair-annihilation luminosity to the jet energy is a
few 10%. Hence Shibata and Taniguchi (2006) concluded that systems composed of a black hole and a hot torus with \( M_d > 0.01 M_\odot \) and an accretion rate \( M \sim 0.1-1 \, M_\odot \, s^{-1} \) are possible candidates for the central engine of SHBs.

Mergers with more equal masses \((q \sim 1)\) produce much less massive, geometrically thin disks and insufficient thermal energy is available. Such mergers cannot power SHBs (Shibata, Taniguchi, & Uryū 2005). Unfortunately, binary neutron stars with small mass ratios \((q < 0.9)\) have not been found so far (Stairs 2004). This is a difficulty for this potential model of SHBs.

### 8.7.4 NS–NS coalescence: the role of the NS magnetic field and its evolution

We have listed and discussed the results of several simulations of binary neutron star mergers, but none has fully taken magnetic fields into account. We have seen the work of Duez et al. (2006) who considered the evolution of magnetized high-mass neutron stars and the importance of the presence of magnetic field. Price and Rosswog (2006) were the first to present global NS merger simulations that follow the evolution of the magnetic field. Their main result is that the existing magnetic field of the neutron star \((10^{12} \, G)\) becomes amplified by several orders of magnitude within the first milliseconds after the merger, long before the collapse to a black hole. A lower limit on the field that can be reached is \(2 \times 10^{15} \, G\). For their simulations they considered two non-spinning magnetized neutron stars with masses \(1.4 \, M_\odot\) and an initial separation of 48 km. Figure 8.17 (their Figure 1) describes the global dynamical evolution of the two neutron stars, which merge into a single object in less than \(~2\, ms\). This new object sheds mass into spiral arms that are subsequently wrapped around the central object to form a hot torus. The magnetic field is amplified dramatically in the shear interface which is formed when the stars come into contact. It is subsequently advected with the matter to cover the surface of the central merger remnant. The amplification leads to a field strength beyond \(2 \times 10^{15} \, G\) on a timescale of only 1 ms. Under these conditions blobs of matter in high field pockets will become buoyant, float up and produce a relativistic blastwave as they break through the surface of the central object (Kluźniak & Ruderman 1998). This could be the variable relativistic outflow that is required to produce a GRB, the duration of the burst being given by the time it takes for the central object to collapse or to consume its rotational energy. A question which is still critical, even in these models based on NS–NS mergers, concerns the baryonic pollution. Relativistic numerical calculations indicate that MHD effects help to have low baryon pollution along the polar axis of the system (Duez et al. 2006, Shibata et al. 2006). At neutrino luminosities \(\sim 10^{52} \, \text{erg s}^{-1}\), however, neutrinos will drive a strong baryonic wind (Duncan, Shapiro, and Wasserman 1986). This material can be dangerous for the emergence of the ultra-relativistic outflow necessary to have a SHB. As the central object is very opaque to neutrinos, the total neutrino luminosity is largely due to the inner shock-heated torus regions. Directly after the merger, the environment has a very low density and rising magnetic bubbles will help to keep the region above the central object relatively clean of baryons, through magnetic pressure. But Price and Rosswog (2006) indicate that it is increasingly difficult to launch relativistic outflows as the luminosity rises and the
Figure 8.17. Snapshots of the coalescence of two magnetized NSs followed between 1.83 ms and 11.34 ms. The color scale on the right gives log(\(B\)) (in gauss). The dimensions of the panel are \(\sim 140\) km from left to right. The stars move gradually toward each other and then merge in a ‘plunging phase’ within about one orbital period (\(\sim 2\) ms, the first two snapshots). This new object sheds mass into spiral arms that are subsequently wrapped around the central object (snapshots 3 to 5) to form a hot torus (last snapshot). The magnetic field is amplified in the shear instability between the stars and subsequently advected with the matter to cover the surface of the central merger remnant (from Price & Rosswog 2006). See also Color section.

continuously braked central object takes longer times to reach buoyancy field strength. The interaction between such magnetic bubbles and a baryonic wind is very complicated and the answer to the question of the production of the relativistic outflow may depend on the details of the merging system. So the question of the pollution along the axis of the system remains an open issue for these merger models too.

To summarize, mergers of neutron stars produce hot, massive accretion disks around spinning black holes and two possible mechanisms can be responsible for the creation of an ultra-relativistic jet. One is the production of \(\nu\bar{\nu}\) pairs in the hot material; \(\nu\bar{\nu}\) annihilation produces a relativistic jet with \(e^+/e^-\) and photons. The viscosity in the disk determines the timescale of the resulting burst, while the total mass accreted into the central black hole may determine the overall energy scale of the burst. For Setiawan, Ruffert, and Janka (2004, 2006) the viscosity is the main parameter defining the overall burst energy (and to a lesser extent the black-hole spin and the disk mass). The other way to extract energy is connected with general relativistic magneto-hydrodynamic effects, which may allow in-falling matter to tap the spin energy of a black hole via the Blandford–Znajek process. In this case
the main parameters determining the energetics are the mass accretion rate and the BH angular momentum. Both mechanisms may make it possible to explain short GRBs, their energy, and their duration.

If NS–NS mergers can produce SHBs, another question has to be answered: are such systems sufficiently numerous to produce the observed rate of SHBs? The estimated rate for these systems ranges from about 4 to $220 \times 10^{-6}$ per year per galaxy (Kalogera et al. 2004a). Rosswog and Price (2006) estimated that this is two orders of magnitude larger than the rate required to explain SHBs; so it seems that there is no problem with the sources, even though beaming has to be taken into account and a fraction of these binary mergers fails to produce SHBs. A word of caution on this positive conclusion is nevertheless recommended because several uncertainties remain: the rate of mergers in the Milky Way is uncertain perhaps by a factor of 10, and the beaming angle and fraction of binary mergers which can produce SHBs are also very uncertain.

In conclusion, with all the reservations we have indicated, NS–NS mergers can be considered as a potentially very interesting source of SHBs. The physical processes at work are not trivial ($\nu\bar{\nu}$ pairs and/or MHD effects to tap the spin energy, the transition from these energy sources to the development of the relativistic outflow) so more work is needed, using full General Relativity to track the evolution and coalescence of magnetized neutron star mergers. In these models an accompanying burst of gravitational waves and neutrinos is expected. If the timescale for the transport of the angular momentum in the torus (by magnetic effects or viscosity) is longer than the emission time of gravitational waves, the black hole is formed before the emission of the GRB (Shibata & Taniguchi 2006), so the emission of gravitational waves could precede the SHB. The detection of these waves with a subsequent detection of a SHB would provide a beautiful and convincing confirmation of this scenario. Finally, while many double neutron star systems are known in the galaxy, other binary systems exist which may be good candidates for the production of GRBs. Let us consider them now.

8.8 MERGERS INVOLVING A BLACK HOLE

8.8.1 Mergers of a black hole and a neutron star

This is the second class of mergers, which received the most attention along with the NS–NS mergers. As early as 1986, Paczyński proposed that GRBs might be at cosmological distances, requiring an energy release corresponding to a supernova explosion. Such a huge power liberated in a very short time would drive a highly relativistic outflow of electron–positron plasma and radiation. The optically thick outflow model proposed by Paczyński (1986; see also Goodman 1986) generates a black-body spectrum, in conflict with observed non-thermal GRB spectra. Paczyński and Goodman proposed solutions to reproduce the observed non-thermal spectra using the fact that the thermal radiation field is at high temperature, $>1$ MeV, and that this temperature will be rapidly variable if the cooling rate is very high. As the
energy involved in cosmological GRBs is comparable to the binding energy of a neutron star, Goodman (1986) had already invoked the merger of two compact objects and specifically of two neutron stars. Paczyński (1991) suggested black hole–neutron star (BH–NS) mergers as possible sources of cosmological GRBs. The neutron star can be tidally disrupted, in which case the debris forms a disk which releases energy on a viscous timescale before being swallowed by the black hole. Mészáros and Rees (1992) suggested, as Goodman (1986) had, the production of a hot photon and $e^+/e^-$ plasma bubble by $\nu\bar{\nu}$ annihilation, and the escape of this ultra-relativistic pair plasma only along the rotation axis. These authors also indicated that the interaction of such an ultra-relativistic jet with external magnetic fields or with matter in a shock would result in the production of a non-thermal spectrum, as observed in GRBs, and would introduce longer radiative timescales, more compatible with GRB light-curves. Baryon contamination was also evoked very early as a dramatic way to degrade the energy of the emergent photons (Rees & Mészáros 1992).

The case of black hole–neutron star mergers was specifically studied by Mochkovitch et al. as early as 1993. Their Figure 1 provides a good, schematic representation of the geometry of such a system, emphasizing the major role of $\nu\bar{\nu} \to e^+/e^-$ conversion within a matter-free funnel formed along the rotation axis (see also Figure 8.11(a)). In this model, neutrino annihilation is the source of the gamma-rays escaping along the funnel. In 1997 Mészáros and Rees discussed the properties of magnetically dominated jet-like outflows from stellar mass black holes surrounded by a debris torus resulting from the disruption of the neutron stars. Magnetic fields had to be extremely high (higher than in typical pulsars) in order to transform the rotational energy into Poynting flux quickly enough. Since the outflow is highly directional, they concluded that jet-like Poynting-dominated (or magnetically driven) flows from tori around black holes can produce even the most intense bursts.

In a series of papers Lee and Kluźniak (1995, 1999a,b) and Kluźniak & Lee (1998), have studied the coalescence of a neutron star with a stellar mass black hole for a range of mass ratios and a stiff equation of state (EOS) of the neutron star (Lee & Kluźniak 1999a) and for a soft EOS (Lee & Kluźniak 1999b). The first objective of these studies was to see if the neutron star was tidally disrupted and if the relativistic fireball could develop with a low baryon contamination. The conclusions of Lee and Kluźniak (1999b) were the following:

- For a soft EOS, and for every mass ratio investigated ($M_{\text{NS}}/M_{\text{BH}} = 1, 0.8, 0.31, 0.1$), the neutron star is entirely disrupted by tidal forces and a dense accretion torus (a few tenths of a solar mass) is formed around the black hole.
- The final accretion rate is between 2 and $6 \, M_{\odot} \, s^{-1}$, so the expected lifetime of the torus is very short, between 40 and 60 ms.
- The rotation axis of the system remains free of matter at a level which allows the production of a relativistic fireball, and the scenario studied can account for the short GRBs.
- The results are dramatically different for a stiff EOS.
With a stiff EOS (Lee & Kluźniak 1999a) an accretion torus appears around the black hole and a baryon-free line of sight along the rotation axis of the binary is present (with less than 10^{-4} M_{\odot} within about 5^\circ) only for high mass ratios q = M_{NS}/M_{BH} = 1. For lower mass ratios (q = 0.8 and q = 0.31), no accretion structure remains around the black hole. All the matter stripped from the neutron star is directly accreted by the black hole, and the neutron star itself is not completely disrupted. Incidentally, this might be the way to produce a low-mass neutron star. Similar conclusions are obtained by Janka et al. (1999) with a physically realistic EOS (Lattimer & Swesty 1991): for mass ratios near unity, the neutron star is destroyed during its first approach to the black hole whereas for low $M_{NS}/M_{BH}$ ratio, the neutron star is disrupted after a few cycles of orbital decay. When the accretion torus can be formed with a mass between 0.3 and 0.7 $M_{\odot}$ the neutrino luminosities which can be obtained reach several 10^{53} \text{erg s}^{-1} during an estimated accretion timescale of \sim 0.1 s. $\nu\bar{\nu}$ annihilation into $e^+/e^-$ pairs with efficiencies of a few percent leads to an energy deposit $E_{\nu\bar{\nu}} < 10^{51}$ erg above the poles of the binary system in a region that contains <10^{-5} $M_{\odot}$ of baryonic matter. The resulting fireball can reach Lorentz factors of around 100. Short GRBs with durations 0.1 to 1 s can be produced at cosmological distances with moderate beaming ($\Omega \sim 0.1$ to 1 sr).

Rosswog, Speith, and Wynn (2004) noted the effect of the rotation of the donor star and reported that a neutron star without spin will more likely settle into an accretion disk outside the Schwarzschild radius. If the donor star is not co-rotating, more material can avoid being fed directly into the Schwarzschild radius. In fact Rosswog, Speith, and Wynn (2004) found that it is very difficult to form a disk around the black hole. Although somewhat easier in the non-rotational case, they found disks with rather low temperature ($\sim 3$ MeV) and moderate densities, which are not promising for powering a GRB by extracting energy via $\nu\bar{\nu}$ annihilation or from the black hole rotation. Moreover, some of the systems investigated do not produce any disk at all. For a stiff EOS a ‘mini neutron star’ can be formed (see also Lee & Kluźniak 1999a), while for a soft EOS the complete tidal disruption of the neutron star is predicted. These conclusions are shared by Lee and Kluźniak (1999a,b) and by Janka et al. (1999).

So, the overall evolution of BH–NS mergers is very sensitive to the mass ratio of the system and to the rotation of the compact objects. The role of the EOS is also crucial: a stiff EOS makes the formation of the disk difficult. The core of the neutron star survives orbiting around the black hole as a mini neutron star for the several 10 ms of the simulation (Rosswog, Speith, & Wynn 2004). If the EOS is soft, the complete tidal disruption of the neutron star is obtained. The conclusions of Rosswog, Speith, and Wynn (2004) are more pessimistic than those of previous authors concerning the possibility of producing GRBs with BH–NS mergers. This seems at odds with the frequently read statement that it is easier to make short GRBs with BH–NS mergers than with NS–NS mergers. The difference between the two systems is the larger energy available for the BH–NS system. But as noted by Popham, Woosley, and Fryer (1999) the efficiency for converting accreted mass energy into neutrinos is highly variable. Greater efficiencies are obtained for low viscosity and high accretion rate. Smaller efficiencies are also obtained if the mass of
the black hole is larger. Of course the first requirement is to have a torus around the black hole. As we have seen, this depends on the rotation of the neutron star, on the ratio $M_{NS}/M_{BH}$, and on the EOS. The physics at suprnuclear densities and the stiffness of the neutron star EOS are studied by Rosswog, Speith, and Wynn (2004) as decisive.

Davies, Levan, and King (2005) considered the ultimate outcome of black hole–neutron star mergers. They insisted on the fact that large mass ratios can dramatically affect the fate of the mergers. They also reanalyzed the mass transfer between the neutron star and the black hole using a semi-analytical approach. First the drag force brings the neutron star and the black hole into contact as the neutron star fills its Roche lobe. Instantaneous mass transfer will follow, with angular momentum transferred back to the neutron star, resulting in an eccentric orbit with the neutron star no longer filling its Roche lobe. The system then spirals in with emission of gravitational radiation. The neutron star again fills its Roche lobe, and a second episode of mass transfer begins. A new kick to the neutron star and the process repeats until the neutron star reaches a mass $<0.2 M_\odot$. At this point the neutron star and the black hole are in contact, the neutron star is shredded and forms a disk with a radius of 150–250 km. This disk is massive enough to ultimately produce a short GRB. If this process occurs the mass of the disk will be about the same at the end, even if the initial mass of the neutron star varies. GRBs produced by such a process might have approximately the same energy (but see the comments of Miller (2005) who considers that such Newtonian simulations are not realistic).

Rosswog (2005) explored a new range of masses for the black holes, from 14 $M_\odot$ to 20 $M_\odot$, never considered before. It is expected that higher masses will lead to more difficulties for building up massive disks. Between 14 and 16 $M_\odot$, disks are formed but they are thin and cold. Beyond 18 $M_\odot$, almost the entire neutron star disappears into the black hole without forming a disk. The conclusion of Rosswog is that none of these systems is a promising GRB progenitor. He discussed one possible exception for very rapidly spinning black holes which might be able to produce a burst because their innermost stable circular orbit (ISCO) is much closer to the black hole, allowing the existence of a hot, high-density disk. In fact as observed by Miller (2005) only low-mass black holes can disrupt a neutron star outside their innermost stable circular orbit. This is due to the fact that the ratio of the tidal radius to the radius of the ISCO decreases with the increasing mass of the black hole. Thus, it is only for low-mass black holes that the tidal disruption occurs outside the ISCO, that an accretion disk can form, and stable mass transfer can occur, otherwise the matter plunges directly into the black hole. But the mass of the black hole is not the only parameter; other effects analyzed by Miller (2005) can enhance the likelihood of a plunge. Among them the gravitational radiation is a significant sink of angular momentum. Angular momentum loss to gravitational radiation starts a plunge orbit well outside the ISCO, causing direct merging even for low-mass black holes if the spin parameter of the black hole is modest. Even if the black hole is spinning rapidly, it may still be difficult to have an accretion disk because the radius of the ISCO is small enough only for black holes which are very close to maximum rotation. So the conclusions of Miller (2005), like those of Rosswog, are pessimistic, leading Miller to contemplate the
possibility that no plausible combination of neutron star and black hole masses, spins and orbital inclinations would result in an accretion disk or stable mass transfer. These early conclusions, however, were not based on fully general relativistic numerical simulations, which are really needed to resolve this problem. So let us now investigate more recent studies of black hole–neutron star mergers with fully general relativistic simulations.

Faber et al. (2006), for instance, undertook such simulations using fully general relativistic hydrodynamics. In the following we call $a_R$ the separation between the black hole and neutron star at which the tidal disruption of the neutron star starts, $q$ ($= M_{\text{NS}}/M_{\text{BH}}$) the mass ratio of the binary, and $C$ ($= M_{\text{NS}}/R_{\text{NS}}$, where $M_{\text{NS}}$ is in solar mass and $R_{\text{NS}}$ is in kilometers) the compactness of the neutron star. The authors recall that for a typical neutron star, with compactness $C = 0.15$, the critical mass ratio at which tidal disruption occurs at $a_{\text{ISCO}}$ is $q \sim 0.24$. For a sufficiently massive black hole (small $q$), $a_R$ is smaller than the radius of the innermost stable circular orbit $a_{\text{ISCO}}$ and the neutron star passes through the ISCO before being disrupted tidally. Faber et al. (2006) considered the evolution of a neutron star with compactness $C = 0.09$ around a black hole, with a binary mass ratio $q = 0.1$. This binary is expected to be tidally disrupted slightly within the ISCO. Figure 8.18 (their Figure 1) shows the evolution of the neutron star projected onto the orbital plane and demonstrates the formation of an accretion disk. While the neutron star spirals in until 98% of its mass lies within the ISCO, the rapid redistribution of angular momentum during tidal disruption leads to the formation of an outwardly directed spiral arm sending some matter back outside the ISCO. For this model, while 0.75 $M_{\text{NS}}$ is accreted by the black hole directly, a disk of mass $\sim 0.12 M_{\text{NS}}$ is formed, and the remaining 0.13 $M_{\text{NS}}$ is completely ejected out of the system. The authors also indicate that all the matter lies in the equatorial plane rather than around the polar axis, as required for a GRB progenitor (Figure 8.18). The matter which is initially cold in the disk is heated via shocks. The inner part of the torus reaches very high temperatures ($T \sim 3–10$ MeV) and surface densities (a few $\times 10^{17}$ g cm$^{-2}$). The neutrino luminosity reaches $10^{34}$ erg s$^{-1}$, leading to an annihilation luminosity of $\sim 10^{49}–10^{50}$ erg s$^{-1}$, comparable to the luminosity obtained in the collapse of a hyper-massive neutron star (HMNS, see Section 8.7). Thus, this hot torus is similar to the one formed by the collapse of an HMNS, and unlike previous pseudo-Newtonian calculations of BH–NS mergers (Janka et al. 1999, Lee & Kluźniak 1999a,b, Rosswog, Speith, & Wynn 2004), Faber et al. (2006) thought that their scenario is able to produce SHBs and can compete with relativistic models involving binary neutron star mergers and the delayed collapse of HMNS. They also estimate that approximately 0.03 $M_{\text{NS}}$ will fall back on the black hole at late times, on a timescale equal to or longer than 1 s, and might produce lower energy bursts at later times. This fall-back accretion might explain the X-ray flares observed in SHBs many seconds after the initial burst, especially if self-gravity leads to the formation of higher-density clumps of material. But as recommended by the authors further simulations are needed to specifically follow the long-term evolution of the accretion torus and of the surrounding material.

These recent views concerning the production of GRBs by BH–NS mergers are less pessimistic. As of this writing it remains unclear if short-burst progenitors should
Figure 8.18. Upper panels: From left to right: initial, early, and intermediate configuration of a NS merging with a black hole of mass $M_{\text{BH}}$ solar masses ($t/M_{\text{BH}} = 1$ corresponds to $7 \times 10^{-5}$ s). This corresponds to case B of Faber et al. (2006), with a mass ratio $q = M_{\text{NS}}/M_{\text{BH}} = 0.1$ and compactness $C = M_{\text{NS}}/R_{\text{NS}} = 0.09 \ M_\odot \text{km}^{-1}$. In this case the neutron star is tidally disrupted within the ISCO (dashed curve) and produces a single mass transfer stream, which wraps around the BH. Later, an accretion disk forms because, while the NS spirals in, the rapid redistribution of angular momentum causes the formation of an outwardly directed spiral arm, sending some matter outside the ISCO. A torus is formed with an initial orbital period $P = 7.4$ ms. Lower panels: Matter configuration at the end of the simulation ($T = 990 \ M_{\text{BH}} \sim 0.07$ s) projected in the equatorial plane (left panel) and in a meridional plane (right panel), showing a hot torus located within $r < 50 \ M_{\text{BH}}$. Bound fluid elements are shown as crosses, unbound elements as points. The calculations account for relativistic dynamics within the ISCO (from Faber et al. 2006).

be attributed in general to NS–NS mergers or if BH–NS mergers could have a significant contribution. This remark leads us to say few words on systems which are not too different from BH–NS systems: black hole–helium star and black hole–white dwarf mergers. Contrary to NS–NS (or BH–NS) mergers these systems will produce long GRBs.
8.8.2 **Black hole–He star mergers**

Fryer and Woosley (1998) considered a compact object (neutron star or black hole) which enters the envelope of its red giant companion (see also Podsiadlowski 2007, Thorne & Żytkow 1977). As the compact object and helium core coalesce, the He core is tidally disrupted into an accretion disk around the compact object. The accretion rate can initially reach $\sim 1 M_\odot \, \text{s}^{-1}$. During the merger process the compact object becomes a black hole, if it was not one already. The end system resembles the final situations already discussed for collapsar, NS–NS, and BH–NS mergers, but with a high angular momentum black hole becoming a Kerr black hole. Due to the high efficiency at which the gravitational energy is released, the neutrino luminosity reaches $\sim 10^{51}$–$10^{52} \text{ erg s}^{-1}$ for $\sim 100 \text{ s}$. With an efficiency of 0.01 to 0.1% to convert this energy into a pair fireball by $\nu \bar{\nu}$ annihilation, long GRBs can be produced with energies in the range $10^{49}$–$10^{51} \text{ erg}$, comparable to the values reported for NS–NS mergers. The rate of such mergers can be larger by an order of magnitude than that of NS–NS and BH–NS binaries. No appreciable gravitational wave signal is expected. Moreover, BH–He mergers occur in star-forming regions at the formation site of the He star.

8.8.3 **Black hole-WD mergers and neutron star-WD mergers**

A white dwarf may also merge with a black hole. The formation scenario parallels those of BH–NS binaries (see for instance Figure 11 from Fryer et al. 1999). Fryer et al. (1999) found that the WDs formed in these binaries will preferentially have large masses ($\sim 1 M_\odot$). These massive WDs are tidally disrupted and form an accretion disk. The accretion into the black hole by way of a disk occurs at a rate of $\sim 0.01$ to $0.07 M_\odot \, \text{s}^{-1}$, and lasts for 10–70 s (Popham, Woosley, & Fryer, 1999). As the accretion disk has a radius much larger than in the case of NS–NS mergers, the accretion timescale is larger and neutrino energy transport less efficient. The energy is much higher for the higher mass WDs, which give accretion rates of $\sim 0.07 M_\odot \, \text{s}^{-1}$. If the black hole rotates, the energy yield can be increased to perhaps $10^{50} \text{ erg}$. In any case, Popham, Woosley, and Fryer (1999) considered that these systems are not likely to be the leading cause of GRBs. Narayan, Piran, and Kumar (2001) also discussed the accretion onto a black hole of various companion stars, among them He stars and WDs. They concluded that if the accretion disk is larger than a few tens or hundreds of Schwarzschild radii the accretion proceeds via convection-dominated accretion flow. In this situation most matter escapes to infinity rather than falling into the black hole. Black hole–WD and black hole–He star binaries fall into this category and the authors concluded that these systems are not good candidates for producing GRBs. This is easily explained: little mass reaches the black hole and as the neutrino annihilation is very sensitive to the accretion rate, the available $\nu \bar{\nu}$ energy is low (see Table 8.2 reproduced from Popham, Woosley, & Fryer (1999), Table 6). If they produce bursts, BH–WD binaries will make long-duration bursts like collapsars and BH–He star mergers.
**Table 8.2.** Approximate results for various evolutionary scenarios. The parameter \( a \) in the third column characterizes the black hole spin (from Popham, Woosley, & Fryer 1999).

<table>
<thead>
<tr>
<th>Model</th>
<th>Duration (s)</th>
<th>( a )</th>
<th>( \log L_{\nu\nu} ) (erg s(^{-1}))</th>
<th>( \log E_{\nu\nu} ) (erg)</th>
<th>Rate(^a) (Myr(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS + NS</td>
<td>0.1</td>
<td>0</td>
<td>50.7</td>
<td>49.7</td>
<td>0.1–5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>51.5</td>
<td>50.5</td>
<td></td>
</tr>
<tr>
<td>NS + BH</td>
<td>0.1</td>
<td>0</td>
<td>(52)(^b)</td>
<td>(51)(^b)</td>
<td>0.1–50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>(52.6)(^b)</td>
<td>(51.6)(^b)</td>
<td></td>
</tr>
<tr>
<td>Collapsar</td>
<td>10</td>
<td>0</td>
<td>48.5</td>
<td>49.5</td>
<td>( \leq 2000 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.95</td>
<td>51.3</td>
<td>52.3</td>
<td></td>
</tr>
<tr>
<td>BH + WD</td>
<td>15–150</td>
<td>0</td>
<td>43.6–47.7</td>
<td>46–49</td>
<td>0.1–50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>44.8–48.8</td>
<td>47–50</td>
<td></td>
</tr>
<tr>
<td>BH + He core</td>
<td>15–500</td>
<td>0</td>
<td>43.6–48.5</td>
<td>46–50</td>
<td>1–1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>44.8–49.6</td>
<td>47–51</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) This rate assumes a supernova rate of 0.02 yr\(^{-1}\).
\(^b\) Rough estimate, because of an optically thick disk.

After the discovery of GRB 060614 by Swift (a 100 s long GRB at low redshift, \( z = 0.125 \), without accompanying supernova; see more details on this interesting burst in Section 4.10.4), King, Olsson, and Davies (2007) proposed explaining such long GRBs without SNe by the merger of a neutron star with a massive WD. GRBs due to NS–WD mergers must be associated with star-formation regions and lie within their host galaxies which can be of any type. Davies, Ritter, and King (2002) estimated the formation rate of such systems in the Galaxy and showed that half of them merge within \( 10^{8} \) yr. King, Olsson, and Davies (2007) concluded that these systems are good candidates if \( M_{\text{WD}}/M_{\text{NS}} > 2/3 \), where \( M_{\text{WD}} \) is the mass of the donor WD and \( M_{\text{NS}} \) the mass of the accreting NS. If \( M_{\text{NS}} = 1.4 \, M_{\odot} \), as it is often found for neutron stars, this condition requires \( M_{\text{WD}} > 0.9 \) and of course less than \( 1.4 \, M_{\odot} \). King, Olsson, and Davies (2007) considered that such systems can make a significant contribution to the GRB rate. Thus, it seems that NS–WD mergers must not be neglected, even though the favorite models are the collapsar model for long bursts and NS–NS and BH–NS mergers for short bursts (see also Levan et al. (2006) for a model involving two white dwarves).

Fryer, Woosley, and Hartmann (1999) discussed the possible scenarios of production of these different systems. They also estimated the daily event rate and the fraction of GRBs within a given distance of their host galaxy. The event rates in the Universe are \( \sim 100 \, \text{day}^{-1} \) for merging neutron stars, \( \sim 450 \, \text{day}^{-1} \) for BH–NS mergers, \( \sim 10^{4} \, \text{day}^{-1} \) for collapsars, \( \sim 10^{3} \, \text{day}^{-1} \) for BH–He star mergers, and \( \sim 20 \, \text{day}^{-1} \) for BH–WD mergers. The uncertainty in these numbers is very large, one or two orders of magnitude. Clearly, BH–NS and NS–NS binaries can contribute to the short GRBs with a significantly larger rate for BH–NS mergers: a factor of 4.5 for Fryer
et al. (1999) or a factor of 10 for Bethe and Brown (1998). Like collapsars, BH–He star and BH–WD mergers contribute to the production of long GRBs, but if the rates are correct, the population of long GRBs is largely dominated by collapsars. Popham, Woosley, and Fryer (1999) considered a black hole of several $M_\odot$ accreting matter from a disk at rates from 0.01 to 10 $M_\odot$ s$^{-1}$. They calculated the energy which can be produced in $\nu \bar{\nu}$ annihilation and through the BZ mechanism for the different models. Solutions are obtained for various values of the accretion rates, of the viscosity parameter and of the black-hole spin rate. Their work allows evaluating the capacity of these models to produce GRBs. Finally, the capacity of merger models to really produce GRBs depends crucially on the stability of the torus, which is function of its thickness, its stratification and of the internal viscous dissipation and neutrino cooling. For a comprehensive discussion of the progenitors of SHBs and of the physics of the merger systems, we recommend the review of Lee and Ramirez-Ruiz (2007).

Let us now address the issue of the progenitors of SHBs from a rather different point of view: can the galactocentric distribution of SHBs be used to discriminate between the two types of progenitors: collapsars vs. mergers?

### 8.9 THE GALACTOCENTRIC DISTRIBUTION OF COMPACT STAR MERGERS

At the beginning of this chapter, we discussed many observational arguments that have been used to prove that the progenitors of most long GRBs are collapsars. What is the situation for short GRBs?

In the case of mergers we have seen that the afterglow is expected to be faint or absent because of the low density of the surrounding medium. This is not, however, a decisive criterion because the absence of optical or radio afterglows has been reported even for long bursts. But a strong clue could be provided by the predicted locations of neutron star mergers with respect to their host galaxies.

First, it seems that all formation channels for neutron star mergers lead to a strong kick given to the system, which can reach a speed of 100 to 200 km s$^{-1}$. To evaluate the distance covered between the birth of the system and the time of the burst, it is critical to know the time needed for the two neutron stars to merge. Bloom et al. (1999) considered three distinct routes for producing NS–NS binaries and found lifetimes of $10^5$–$10^9$ yr. Later Belczynski and co-workers (Belczynski, Bulik, & Kalogera 2002, Belczynski, Bulik, & Kluźniak 2002, Belczynski, Bulik, & Rudak 2002, Belczynski, Kalogera, & Bulik 2002) reported that a large fraction of mergers could occur on much shorter timescales, typically $10^5$–$10^6$ yr. They considered a neutron star in a tight binary with a helium star that will produce the second neutron star when it explodes as a Type II supernova. They assumed that the mass transfer is dynamically unstable and leads to a common envelope phase.

---

*The first observations of SHB afterglows are discussed in Section 8.10 below and in Section 4.11.*
If the timescale is long, typically $10^8$ years, one crudely obtains a typical travel distance of $\sim 30$ kpc, assuming that the binary system has gained a space velocity of 100 to 300 km s$^{-1}$ from the two supernova events. This is about $\sim 6''$ at $z = 3$, a separation which is quite observable. Of course, if the timescale is much shorter, as suggested by Belczynski et al., the merger will be within the host galaxy and next to the region of star formation. It is thus crucial to evaluate the typical lifetime of the binary system before the merger, and to understand the evolution leading to the system of binary compact stars.

Considering a population of NS–NS and BH–NS mergers, Perna and Belczynski (2002) computed the distribution of offsets from the host galaxies and of the density of the surrounding environment, from which they inferred the characteristics of the expected GRB afterglows. The introduction of the new class of short-lived neutron star binaries, which dominates the merger rate (80% of the population of NS–NS mergers), clearly changes the results. The histogram of projected distances has a maximum at $\sim 1$ kpc for NS–NS mergers and at a few kiloparsecs for BH–NS mergers (Figure 2 of Perna & Belczynski 2002). Hence, neutron star–neutron star mergers are expected to occur mostly within their host galaxies. Figure 4 of the same authors gives the ambient densities in which the mergers occur. NS–NS mergers probe typical interstellar medium densities while a substantial fraction of BH–NS events occur in a very low-density environment (Perna & Belczynski 2002).

The scenario proposed by Belczynski, Bulik, and Kalogera (2002) for NS–NS mergers was disputed by Rosswog, Ramirez-Ruiz, and Davies (2003), and revisited by Ivanova et al. (2003) considering an alternative for the treatment of the mass transfer. They found that the majority of double neutron star progenitors evolve through a stable mass transfer phase rather than through a common envelope phase and that the fraction of very short-lived (<1 Myr) NS–NS binaries has to be decreased. For the model examined by Ivanova et al. (2003) the fraction of short-lived NS–NS binaries is 15%, instead of 50% in Belczynski, Bulik, and Rudak (2002). While in Belczynski, Kalogera, and Bulik (2002) the empirical coalescence rate was increased by a factor of 2.5 to account for short-lived NS–NS binaries, this factor was only 1.3 in Ivanova et al. (2003). Belczynski et al. (2006) performed an updated analysis of compact binary mergers (NS–NS and BH–NS). They confirmed the two distinct populations: a classical population of long-lived systems with merger times of 100–15 000 Myr and a population consisting of tight, short-lived systems with merger times of 0.001–0.2 Myr. For NS–NS systems, the proportions of short- and long-lived systems are roughly similar, while for BH–NS systems only 20% are short-lived. The authors note that the four known Galactic NS-NS systems are long-lived classical systems.

The lifetime of double neutron star systems has implications for the typical galactocentric distances of the mergers. The implications are small for large galaxies like the Milky Way whose deep gravitational potential prevents the mergers from escaping far from the galaxy borders; in those galaxies the mergers occur inside the host. In small, less massive galaxies, the expected distribution of distances to the host galaxy depends more strongly on the lifetime of the NS–NS binaries before they merge. In this case, the new calculations by Ivanova et al. (2003) indicate that up to
20% to 30% of the mergers may take place outside such galaxies (this fraction was only 5–15% in the work of Belczynski, Bulik, & Kalogera (2002)). In any case, this percentage is significantly lower than previous estimates, which predicted that 50% to 80% (depending on the mass of the host galaxy) of double neutron star systems might escape their host galaxies (Bloom et al. 1999, Bulik 1999). Figure 8.19 (Figure 8 of Ivanova et al. 2003) gives the cumulative distribution of the galactocentric distance of NS–NS mergers in two types of galaxies. This figure shows the strong influence of the gravitational potential of the galaxy. Rosswog, Ramirez-Ruiz, and Davies (2003) have also performed a calculation of the distribution of galactocentric distance of NS–NS mergers for three different galactic potentials. They concluded that the median distance is about 3 kpc. Recently, Belczynski et al. (2006) found that most double compact binaries (>80%) formed within large galaxies merge within their host. For small galaxies this fraction is 20%, 50%, and 70% for elliptical, spiral and starburst hosts respectively. Their conclusion is that SHB progenitors would have to originate from a diverse population of compact objects formed in both old and young stellar environments. Independently of the updated population synthesis code calculations proposed by Belczynski and co-workers, it is difficult to say whether the majority of NS–NS or BH–NS mergers should be within or outside their host galaxy because the conclusions are dramatically dependent on the characteristics of the host galaxy. This has been clearly illustrated by Belczynski, Bulik, and Rudak (2002) for different mergers: BH–NS, BH–WD, NS–NS, NS–He star, NS–WD (their Figures 8, 9, 10, 11, and 12).

Finally, the type of the host galaxy can be also used to discriminate between the two types of progenitors since collapsars must take place in actively star-forming
galaxies while mergers have less connection with such galaxies. For instance, the progenitors of GRBs occurring in elliptical galaxies cannot be collapsars but must be some kind of long-lived binary systems of compact objects. Another remarkable clue would be the nearly coincident detection of a signal in gravitational waves, which is the hallmark of binary merger events (see Chapter 9). In the next section, we put these theoretical studies in perspective with the observations of SHB and their host galaxies.

8.10 SOME RECENT RESULTS ON SHORT GRBS AND THE MERGER MODEL

The first X-ray afterglow of a short GRB (duration \(\sim 40\) ms) was detected by Swift for GRB 050509b (Gehrels et al. 2005). Thanks to Swift’s XRT, the X-ray afterglow was observed very early, 62 s after the burst. The X-ray emission for this burst was faint, the weakest afterglow of any of the 15 GRBs that the XRT had promptly observed at that time, and it faded below the XRT detection limit within only a few hours. No optical afterglow was detected down to stringent limits (\(R\) magnitude >25 at 25 h, Cenko et al. 2005). Deep optical images with the Very Large Telescope (VLT) at ESO were obtained from 2 days after the GRB trigger and for 3 weeks. No variable object inside the small XRT error circle was detected down to \(V = 26.5\) and \(R = 25.1\) (Hjorth et al. 2005a). The center of the XRT error box lies only 9.8′′ from the center of a large \(E_1\) elliptical galaxy situated at redshift \(z = 0.225\). This galaxy is a giant and bright elliptical galaxy, with no on-going stellar formation. Gehrels et al. (2005) noted that this was the first GRB, of \(\sim 80\) with accurate optical/X-ray localizations, to be near a bright elliptical galaxy. Such galaxies are highly deficient in short-lived massive stars, and they are very different from the hosts of long GRBs, which are typically sub-luminous and blue, comparable to the faint blue star-forming galaxies found at high redshift (Le Floc’h et al. 2003). If this association of a short GRB with a luminous elliptical galaxy having no evidence for on-going star formation (Bloom et al. 2006) is correct, this is certainly a strong argument in support of a merger origin of short, hard GRBs (Gehrels et al. 2005, Bloom et al. 2006).

Two months after this exciting finding, the first optical afterglow of a short GRB was discovered thanks to HETE-2. GRB 050709 was short (\(\sim 70\) ms), but it presented a long, soft emission which was detected by the X-ray detectors of HETE-2 (WXM and SXC) from 25 s to 130 s after the initial hard pulse (Villasenor et al. 2005). Similar behavior had been previously observed for SHBs with the BATSE and KONUS instruments. It was reported by Connaughton et al. (2002) as a common feature of many long bursts and perhaps also of the shorter bursts, at a lower level. Villasenor et al. (2005) proposed that this emission was the beginning of the afterglow, its time history and spectrum being consistent with those expected for the X-ray afterglow (it is, however, more generally accepted now that this late-time X-ray emission is comparable to the X-ray flares following long GRBs; see Section 4.9.2). The late-time X-ray afterglow was discovered with Chandra (Fox et al. 2005). Using the Chandra position a marginal detection of the X-ray afterglow was obtained in
observations made by Swift’s XRT, 1.6 days after the burst. The X-ray observations suggested strong flaring activity which lasted until 16 days after the burst. Fox et al. (2005) attributed this flaring activity to flares arising from the ongoing activity of the central engine analogous to the bright X-ray flares seen in several long-duration GRBs (see Section 4.9). Extensive ground-based follow-up campaigns at radio, optical, and near-infrared wavelengths were also conducted, leading to the identification of the optical afterglow and of the host galaxy of GRB 050709 (Hjorth et al. 2005b). This host was a late-type, star-forming, dwarf blue galaxy at redshift \( z = 0.160 \). HST observations identified a single bright fading point-like source, which was unambiguously the optical afterglow (Fox et al. 2005). It faded with a decay index similar to that of long GRBs. So for this burst the afterglow exists, and a break in the decay was reported by Fox et al. (2005), but not confirmed by Watson et al. (2006). The host galaxy is a typical star-forming irregular galaxy. The burst was offset by 3.8 kpc from the brightest central region of the host galaxy (Fox et al. 2005, Hjorth et al. 2005b); this is among the largest offsets observed for long GRBs (Bloom et al. 2002). The upper limits on the optical emission between 7 and 20 days after the burst ruled out any type Ic supernova brighter than 1–2% of SN 1998bw (Hjorth et al. 2005b). Even though for this burst the progenitor was in a star-forming dwarf galaxy, the absence of a SN and the offset of the burst from the host galaxy are characteristics that are quite distinct from those of long GRBs. Hence, GRB 050709 and GRB 050509B suggested different origins for the short GRBs, although for different reasons.

One week later, another short burst GRB050724, was localized by Swift (Berger et al. 2005), giving the opportunity to observe the radio afterglow of a short GRB for the first time. Here again soft X-ray emission was seen with BAT for >150 s after the initial hard spike. Figure 8.20 (Figure 3 of Barthelmy et al. 2005) shows the evolution of the X-ray afterglow, combining data from BAT, the XRT, and Chandra. For this burst too flaring activity is visible, especially at \( \sim 2 \times 10^4 \) s, and it was explained by late internal shocks (Barthelmy et al. 2005, Campana et al. 2006b, Grupe et al. 2006). These flares are thus a common feature of short and long GRBs. A fading optical afterglow was also observed (Gal-Yam et al. 2005), located approximately 1” south of the center of (but within) a bright elliptical galaxy at redshift \( z = 0.258 \). The projected offset from the center of the galaxy is \( \sim 4 \) kpc and the identification of the elliptical galaxy as the host galaxy is considered quite secure (Barthelmy et al. 2005). From the modeling of the afterglow Panaitescu (2006) deduced a density of the surrounding medium in the range \( n = 10^{-1} \) to \( 10^{-3} \) cm\(^{-3} \), larger than that of a galactic halo or of the intergalactic medium and consistent with the small offset. Again such large elliptical galaxies are very interesting sites for old compact binary systems and promising locations for NS–NS or BH–NS mergers. Such binary progenitors are generally believed to have a coalescence time of \( \sim 1 \) Gyr, but the small offset suggests that the kick velocity imparted to the system was probably too small to unbind it from the host (Berger et al. 2005) or that the merging time was shorter than 1 Gyr as suggested by Belczynski, Bulik, and Kalogera (2002). The multi-wavelength modeling of the afterglow, from radio to X-rays, allowed Panaitescu (2005) to derive a jet half-angle larger than 8° (see also Berger et al. 2005, who give 8.5°), wider than for long GRBs,
which have a median value of 5°. This value is significantly lower than the one derived by Grupe et al. (2006) who found no break in X-ray observations of the afterglow with Swift and Chandra ranging from 100 s to 3 weeks after the burst. This lack of jet break places a lower limit of 25° on the jet opening angle and gives a beaming corrected energy of $4 \times 10^{49}$ erg at least. If such a significant difference between the jet opening angle of SHBs and long GRBs is confirmed, this would reduce the gap in the total energy budget of short and long GRBs implied by isotropic energies, which differ by several orders of magnitude.

With these three examples it appears that the progenitors of short GRBs are created in all types of galaxies, suggesting a correspondingly wide distribution of delays between the formation and the explosion of the progenitors (Prochaska et al. 2006). Fox et al. (2005) have summarized the physical properties of short, hard bursts and their host galaxies. They compared four bursts: the three that we have just discussed plus GRB 050813. Their redshifts are low, $z < 1$, with three of them having $z < 0.25$. Their isotropic equivalent energies are lower than a few $10^{50}$ ergs, i.e. significantly less than the values associated with long GRBs ($10^{52}$ to $10^{54}$ erg). Thus, an energy release that is at least two or three orders of magnitude below the energy released by long GRBs seems to be the rule (but see the previous remark on the collimation angle). The case of GRB 050813 is very interesting and symptomatic of the appearance of new questions about the distances of SHBs (due to the fact that the progenitor is allowed to migrate far from its birthplace). This short burst had an X-ray afterglow, but no optical or radio afterglow. The initial X-ray position led to two
possible host galaxies being proposed at \( z = 0.72 \) (Berger et al. 2005, Foley, Bloom, & Chen 2005). However, a revised XRT position seemed to exclude these galaxies, but contained a fainter galaxy at \( z = 1.8 \) (Berger 2006). This redshift was significantly higher than those of the previous short GRBs. A detailed analysis of the observations of this burst was performed by Ferrero et al. (2007), showing that both redshifts are indeed possible. Similar issues were raised by GRB 061201 which had well-identified X-ray and optical afterglows and had no detectable host down to magnitude \( R \sim 25 \). This burst was studied by Stratta et al. (2007). In the absence of a priori information on the type of host, it could be associated with about the same probability with a cluster of galaxies at \( z = 0.085 \) (with an offset of \( 8.5' \), corresponding to a physical distance of \( \sim 800 \text{ kpc} \); Berger 2007b), with a star-forming galaxy at \( z = 0.111 \) (with an offset of \( 17'' \), corresponding to \( \sim 34 \text{ kpc} \) from the center of the galaxy), or with an invisible galaxy (too faint to be detected) at \( z > 1 \). The case of this burst is even more intriguing than GRB 050813, since the clear optical afterglow leaves no doubt on its position on the sky. Unfortunately, the spectrum of the optical afterglow showed no line which could help in determining the redshift of the burst. We will see below that more recent SHB observations suggest that the redshifts of SHBs detected with Swift might extend to \( z \sim 1 \) and beyond.

Prochaska et al. (2006) presented the spectra of the host galaxies obtained with the Gemini-North and Keck telescopes for the first three SHBs. As we have seen, two of these bursts (GRB 050509B and GRB 050724) were found to be associated with old and massive galaxies with no current (<0.1 \( M_{\odot} \text{ yr}^{-1} \)) or recent star formation. The third (GRB 050709) occurred within a dwarf galaxy with a star-formation rate exceeding 0.3 \( M_{\odot} \text{ yr}^{-1} \). The authors concluded that these observations, of course too scarce, support a merger origin of SHBs and suggest that the progenitors of SHBs are created in all types of galaxies with a wide distribution of delays between formation and explosion, like type Ia supernovae. Using this same sample, and GRB 051221, Guetta and Piran (2006) reported that this sample is incompatible with a SHB population which would follow the star-formation rate, although it is compatible with a population having a distribution of delay times after the SFR, in agreement with the conclusion of Prochaska et al. (2006). These conclusions support the idea that binary neutron star mergers are possible progenitors of SHBs. The sample used in these studies is, however, too small to allow robust conclusions. At the end of this famous year 2005, another short GRB was detected by Swift: GRB 051221A. The optical, X-ray and radio observations of the bright afterglow were reported by Soderberg et al. (2006a). These multi-wavelength observations allowed detailed studies of the afterglow. The redshift of the burst was determined to be \( z = 0.5464 \). The isotropic equivalent prompt energy was about \( 8.4 \times 10^{51} \text{ erg} \). An observed jet break at \( t \sim 5 \text{ days} \) allowed determining an opening angle \( \theta_j \sim 7' \), comparable with the values obtained for GRB 050709 and GRB 050724 (see also Burrows et al. 2006). The beaming-corrected energy was \( \sim 2 \times 10^{49} \text{ erg} \), comparable with previous short GRBs. The host galaxy was actively forming stars but at the same time it enclosed a significant population of old stars (\( \sim 1 \text{ Gyr} \)) and it had a near solar metallicity. No bright supernova emission was observed and the circum-burst density was low (\( n \sim 10^{-3} \text{ cm}^{-3} \)). These observations continue to support the idea that short
GRBs are not related to the deaths of massive stars but rather to the coalescence of compact objects (Soderberg et al. 2006a).

In spite of these uncertainties and of the diversity of the properties of short GRBs, we can tentatively propose the following characteristics deduced from the analysis of the very small sample of the first five well-localized short GRBs:

- The afterglows of SHBs seem fainter than those of long GRBs.
- The external shock model is also relevant for the afterglows of SHBs.
- SHBs apparently occur in host galaxies of all types, as do SNe Ia. The presence of early-type hosts argues for a delay greater than 1 Gyr between the formation of the progenitor and the burst (Bloom & Prochaska 2006).
- The population is heavily skewed to low redshifts compared with the population of long GRBs ($\langle z \rangle \sim 2.8$; Jakobsson et al. 2006). For the first five SHBs, $\langle z \rangle \sim 0.38$ (assuming that GRB 050813 took place at $z = 0.72$), but a higher mean redshift is still possible (see the discussion about the redshift distribution in Bloom & Prochaska (2006)), although in any case smaller than that of long GRBs. Of course more redshift determinations of SHBs are needed. A recent analysis of Berger et al. (2007) presents deep optical observations of nine well-localized SHBs (four with sub-arcsecond positions and five localized to $6''$ by the XRT). They find that eight of the nine bursts appear to be associated with galaxies fainter than $R \sim 23$ (and the remaining with a galaxy with $R = 21$). This contrasts with previous short GRBs that were associated with galaxies brighter than $R \sim 22$. The spectroscopic redshifts of the four brightest hosts in their sample are: $z = 0.41, 0.44, 0.83$, and 1.13, comparable with or larger than the redshifts of the first five SHBs that we have already mentioned. This suggests that the remaining five SHBs, with faint hosts, would lie at higher redshifts (see Figure 8.21; from Berger et al. 2007). The most conservative limit given by Berger et al. (2007) is that at least half of this new sample resides at $z > 0.7$, suggesting that 35% to 65% of all SHBs originate at higher redshifts than previously reported. Of course, if this analysis is confirmed, it will lead to a significant revision of the conclusions based on the host galaxy and redshift determinations made by Swift and HETE-2 in the famous year 2005. Another consequence of this new average distance of SHBs is a revised value of the isotropic energy, which was between $10^{48}$–$10^{49}$ erg with the first identified SHBs at low redshift. Now it could be significantly larger, between $10^{50}$ and $10^{52}$ erg. These results also have an impact on the time delay prior to the GRB explosion: a high redshift population implies that all progenitors cannot be several gigayears old, and that the typical ages would be shorter than previously reported.
- The location within the host galaxy is also an important issue. For the first SHB, GRB 050509B, a relatively large offset from the host galaxy is observed: $39 \pm 13$ kpc in projection, but this offset is less than 10 kpc for GRB 050709, GRB 050724, and GRB 051221 (Table 2 of Bloom & Prochaska 2006). For long GRBs this offset does not exceed 10 kpc. Here again, the offset distribution needs a larger sample to be confirmed. Nevertheless, what appears relatively secure is the conclusion of Bloom and Prochaska (2006): the progenitors of SHBs cannot
all have both large systematic kicks (>100 km s\(^{-1}\)) at formation and large delays from starburst (>1 Gyr).

- The SHBs tend to be located in low-density environments with densities clustering around 10\(^{-2}\)–10\(^{-3}\) cm\(^{-3}\) (Soderberg et al. 2006a). This is also the conclusion of La Parola et al. (2006) who found \(n < 3 \times 10^{-3}\) cm\(^{-3}\) for GRB 051210, but these authors stressed the uncertainties linked to this estimate. In any case, if this result is confirmed it would give a mean density of the environment lower than that inferred from long GRBs and in agreement with low values expected in the merger models.

- Flaring activity, which has been observed for several SHBs, appears to be a common property of short and long GRBs. This is interesting and intriguing because flares, which may occur several tens of seconds after the GRB, seem more difficult to produce in the context of merger models. If GRBs are powered by hypercritical accretion flow onto a BH the dynamical timescale is only milliseconds, while the viscous timescale is at most a few seconds, still dramatically shorter than the delay of X-ray flares after the GRB. The hydrodynamical simulations of Davies, Levan, and Singh (2005), for instance, suggested that the central engine activity of merger events cannot last more than a few seconds. Nevertheless, Perna, Armitage, and Zhang (2006) proposed a common origin for flares in short and long GRBs. They suggested that gravitational instabilities lead to the fragmentation of the rapidly accreting accretion disk that forms in the collapsars and merger models (with different masses), creating blobs of material whose in-fall onto the central object may produce the observed flares even at late time. In Section 7.4 we have indicated other ways to produce these X-ray flares, for instance magnetically powered events, which may be able to offer longer periods of activity.
If we consider the sample of 16 short-duration GRBs analysed by Berger (2007a) and the comprehensive review of Nakar (2007), we can add some important properties which are not directly connected with the progenitors but which are very useful for characterizing the SHB populations:

- The majority of SHBs (80%) follow a linear correlation between $E_\gamma$ and $E_k$ (the kinetic energy of the ejecta, measured by the X-ray luminosity at one day), indicating a narrow distribution of the gamma-ray efficiency, $\varepsilon_\gamma$ ($\sim 0.1$), similar to the values for long GRBs.
- The density of the surrounding medium is generally larger than $n \sim 0.05 \text{ cm}^{-3}$.
- The isotropic-equivalent energies span $10^{48} - 10^{52} \text{ erg}$, similar in width to the distribution for long GRBs.
- Possible beaming seems to indicate that the true energy release has a narrow distribution, as for long GRBs, but with significantly lower energy: $\sim 10^{48} - 10^{49} \text{ erg}$.

A word of caution is necessary to accompany this list of SHB properties because the current sample is too small to draw strong conclusions. Nevertheless, we can say for sure that the collapsar model cannot be responsible for the short GRBs, whereas the present reported characteristics of short GRBs and their afterglows can be accommodated by merger models. Following the association of long GRBs with type Ic supernovae, it may be tempting to relate SHBs to binary mergers: spun-down hypermassive neutron stars (HMNS) and/or accretion induced collapse of a neutron star. In that case, the host galaxies of SHBs would also harbour type Ia SNe whose progenitors belong to an old and evolved population, involving a white dwarf in an interacting binary system (van Dyk 1992). This idea led Zhang et al. (2007) to propose considering two classes of GRBs: type I and type II. Type I GRBs would be associated with old stellar populations similar to type Ia SNe with compact star mergers as favorite progenitors. Type II GRBs would be associated with young stellar populations, like type II and type Ibc SNe with core-collapse supernovae as favorite progenitors. In both cases the ultimate system consists of a dense torus of matter around a black hole and the mass of the torus and the environment make the difference between short and long GRBs. For long GRBs the more massive torus fed by an external reservoir of stellar matter would explain their larger and longer energy generation. In their case, the fall-back of material associated with the SN explosion may play an essential role (e.g. Janiuk et al. 2004).

This idealized picture may, however, be too simple, and the population of SHBs could be more diverse than we believe. The detection by RHESSI and Swift of a giant flare from the galactic magnetar SGR1806–20 (Hurley et al. 2005) has initiated discussions about the fraction of similar events (from sources in the local Virgo cluster of galaxies) present in the current sample of short, hard GRBs. There is, however, a consensus that the majority (more than 80%) of SHBs are not due to giant flares from extragalactic magnetars. After all is said, it seems that mergers of compact stars are really needed to explain a large fraction of SHBs, justifying all the theoretical work that has been presented in the previous sections.
8.11 CONCLUSION

The general conclusion of this chapter is that at least two main classes of progenitors have been identified for long and short GRBs. They involve the collapse of massive stars (a single Wolf–Rayet star or a system of two WRs) or the coalescence of two compact objects (two neutron stars or a black hole and a neutron star). Both types of progenitors lead to similar systems: a black hole with a debris torus, with few solar masses for the black hole. In the first class of progenitors the disk mass may be heavier (several $M_\odot$), while in the case of coalescence the disk might be less massive (a few $\times 10^{-1} M_\odot$). The overall energetics from these progenitors does not differ by more than one order of magnitude (Mészáros, Rees, & Wijers 1999). The two types of progenitors may differ more significantly by the amount of rotational energy contained in the black hole. For both types of progenitors two large reservoirs of energy are available: the binding energy of the orbiting torus and the spin energy of the black hole (Mészáros & Rees 1997). The first can provide up to 40% of the rest mass energy of the disk (for a disk around a maximally rotating black hole), while the second can provide up to 29% of the rest mass of the black hole.

We have presented three processes to extract the energy of the system:

(a) Neutrino annihilation ($\nu + \bar{\nu} \to e^+ + e^-$), which can tap the thermal energy released by viscous dissipation in the torus.
(b) Extraction of the rotational energy of the torus, by the dissipation of magnetic fields generated by the differential rotation in the torus.
(c) Extraction of energy from the black hole itself by the Blandford and Znajek process (1977).

The energy available can be larger than that contained in the orbiting debris because the black hole is rapidly spinning and more massive than the torus. As noted by Mészáros, Rees, and Wijers (1999), the total energy difference between the various models is at most a factor of 20, whereas in the rotationally (BZ) powered cases they differ by at most a factor of few. If the jets are powered by any of these mechanisms, the isotropic energy which can be extracted is sufficient to avoid extreme beaming. So, from the energetic point of view, collapsars and mergers of compact objects are possible progenitors. But how can we recognize them?

There are solid arguments supporting the collapsar model as the explanation of long GRBs. The association of a few nearby long GRBs with supernovae is certainly the more convincing one. Even though the collapsar model has been extensively studied and compared with observations, several points are not clear. They concern, for instance, the nature of the progenitor star: is it isolated or in a binary system; the origin of the jet and its development through the stellar envelope; the physical processes powering the fireball: neutrino annihilations, Poynting flux, or both. But, to be optimistic, we can say that the framework of the model is well determined. We have also mentioned models involving magnetars with very large magnetic fields and very high speeds of rotation. Pulsars exist and magnetars, even though they are much rarer, also seem to exist. In this case the huge magnetic fields involved can
provide the energy source of GRBs with no problem, so collapsars may not be the only progenitors of long GRBs.

The problem of short GRBs is more complex. The main reason is the small number of short GRBs with an observed afterglow. Little information is available on the afterglows, which was the most important source of progress in the study of long GRBs; more data are clearly needed. We nevertheless note that extensive theoretical studies have demonstrated the potential of compact binary mergers, especially NS–NS mergers. Finally, it is worth mentioning that other models (collapsar or giant magnetar flare) face problems (the detection of SHBs in elliptical galaxies for collapsars and the luminosity of giant magnetar flares) that prevent them from explaining the observed properties of SHBs in most, but not all cases, reinforcing the predicted crucial role of compact binary mergers in the explanation of SHBs.

8.12 REFERENCES


9

Perspectives in gamma-ray burst science

This chapter is organized into three parts:

- The first part discusses GRBs in the context of the knowledge of the early Universe. We show that they can be a formidable tool to study the first generation of stars, the metal-enrichment of galaxies, and the young Universe at the epoch of re-ionization which ended the period of the dark ages. Moreover, if GRBs can be used as standard candles with standardized energies, their detection at high redshift would be a potential tool to extend beyond $z = 2$ the remarkable results already obtained by the observations of type Ia SNe. Hence GRBs would participate in the determination of the cosmological parameters.

- The second part concerns the emissions which may accompany GRBs and their afterglows. These emissions include very high-energy photons and non-electromagnetic emissions: high-energy neutrinos, cosmic rays (CRs) and ultra-high energy CRs (UHECRs), and gravitational waves (GWs). GWs are directly connected with the nature of the progenitors, while the high-energy photons, the neutrinos, and UHECRs may be related to the Fermi acceleration in shocks developing in the GRB outflow. These shocks explain the electron acceleration and thus the GRB with its prompt emission and its afterglow. But if electrons are accelerated, protons can also be accelerated to observer frame energies up to $10^{20}$ eV (Waxman 1995, 2004). These emissions, if they are observed in connection with GRBs, would be of the highest interest in constraining parameters of the shock model.

- The third part is really our conclusion. It includes a non-exhaustive list of problems, which have already been indicated in previous chapters. They concern mainly the standard shock scenario, which remains the most amenable for testing and quantitative modeling, in spite of its difficulties (Zhang & Mészáros 2004). We will also say few words on upcoming space missions which will complete the exceptional work of Swift and on the strategy of ground-based experiments.
Early and long follow-up multi-wavelength observations are critical for studying the afterglow as completely as possible, particularly in the IR band. Thanks to IR observations it will be possible to follow high-redshift GRBs, with their potentially very exciting consequences. Obviously further theoretical investigations will be mandatory in parallel with observations in a large variety of domains.

Let us us start with the use of GRBs as potential probes for the early Universe.

**GRBs and the early Universe**

### 9.1 COSMIC RULERS

One field where GRBs might become an interesting tool for observational cosmology is the description of the geometry of the Universe. Several groups have proposed using GRBs as cosmic rulers to set new constraints on the cosmological parameters $\Omega_M$ and $\Omega_A$, where $\Omega_M$ and $\Omega_A$ are the present-day density parameters of the matter and dark energy components of the Universe respectively.

Observations suggest that the Universe is flat, with $\Omega_{\text{tot}} \sim 1$ ($0.98 < \Omega_{\text{tot}} < 1.08$). $\Omega_{\text{tot}} = \rho_{\text{tot}} / \rho_c$ is the ratio of the total energy density, $\rho_{\text{tot}}$ to the critical energy density, $\rho_c = 3H_0^2/8\pi G$, where $H_0$ is the Hubble constant, the rate of expansion of the Universe at present, and $G$ is the gravitational constant. The fractional contribution of the various components of the Universe is defined by $\Omega_i = \rho_i / \rho_c$ with $i$ denoting either baryons (B), dark matter (DM, possibly made of weakly interacting massive particles), radiation (R), and dark energy (DE). The values of $\Omega_i$ are given, for instance, by Padmanabhan (2006a): $\Omega_R \sim 5 \times 10^{-5}$, $\Omega_B \sim 0.04$, $\Omega_{DM} \sim 0.26$, and $\Omega_{DE}$ called also $\Omega_A \sim 0.7$. With $\Omega_{\text{tot}} \sim 1$ and $\rho \sim \rho_c$, we are in a flat (Euclidean) Universe (inflation models predict nearly exact flatness), which is largely non-baryonic, with a dark energy component (DE) which is certainly the strangest one. Padmanabhan (2006a) considers this dark energy to be the mystery of the millennium and a major challenge in theoretical physics. Indeed, the investigation of dark energy and its nature has become a major issue, especially after the measurement of the Hubble Diagram of type Ia supernovae, which has revolutionized cosmology, showing that the current expansion of the Universe is accelerating (Perlmutter et al. 1999, Barris et al. 2004, Riess et al. 2004, Astier et al. 2006), thus providing convincing evidence for the existence of dark energy.

At high redshift the Universe is radiation-dominated. As the Universe expands there is a time at which the matter energy density becomes comparable with the radiation energy density. At lower redshift, matter will dominate over radiation and ultimately the dark energy density will dominate over non-relativistic matter because, as the Universe expands, the matter energy density falls as the third power of the scale while the energy density contributed by the cosmological constant does not change.\(^1\) For $\Omega_{DE} = 0.7$ and a non-relativistic matter density $\Omega_{NR} = \Omega_B + \Omega_{DM} = 0.3$

\(^1\) If dark energy is indeed a cosmological constant with no variation with redshift (Padmanabhan 2006b).

---

478 Perspectives in gamma-ray burst science | Ch. 9
(\(\Omega_B \sim 0.04\) and \(\Omega_{DM} \sim 0.26\)), the transition at which the dark energy density dominates over non-relativistic matter occurs at \(z_{DE} \sim 0.33\) (with \(1 + z_{DE} = (\Omega_{DE}/\Omega_{NR})^{1/3}\); Padmanabhan 2006a). Assuming a flat Universe with \(\Omega_M = 0.28\), Perlmutter et al. (1999) found \(z \sim 0.37\) for the cross-over between mass-dominated and dark-matter-dominated energy density. This corresponds to a cross-over between deceleration and acceleration occurring at \(z = 0.73\), over \(6 \times 10^9\) yr ago. Riess et al. (2004) found \(z = 0.46 \pm 0.13\) for this transition.

Dark energy with negative pressure is the most generally accepted explanation for this acceleration. Such an exotic component has been disputed, see for example Leith, Ng, and Wiltshire (2008). See also references in Rapetti et al. (2007) or Spergel et al. (2007) and the discussion in Padmanabhan (2006a) for alternative cosmological models. But strong indications of a dark energy component existed before the SN data were known. Already in 1975 Gunn and Tinsley (1975) had suggested that the most plausible cosmological models have a positive cosmological constant. Efstathiou, Sutherland, and Maddox (1990) discussed observations which could be accommodated in a flat cosmology, in which as much as 80% of the critical density was provided by a positive cosmological constant which is dynamically equivalent to endowing the vacuum with a non-zero energy density. As another example, Ostriker and Steinhardt (1995) also evoked a substantial contribution to the energy density from the vacuum itself (a positive cosmological constant) with the critical question: how can one explain its non-zero value?

The presence of this new dark energy component, counteracting the self-attraction of matter, would cause the expansion of the Universe to accelerate if it has a sufficiently negative pressure. The new observations of the distance–redshift relation of type Ia SNe, completed by other recent observations of the spectrum of cosmic microwave background (CMB) anisotropies, of the distance–redshift relation of X-ray clusters of galaxies, and of the amplitude of matter fluctuations from clusters of galaxies, have led to the definition of the concordance \(\Lambda\) cold dark matter (\(\Lambda\) CDM) cosmology (see for instance Rapetti et al. (2007) for references to the observations which have led to this model). In this model the Universe is geometrically flat, homogeneous, and isotropic on large scales. It contains only \(\sim 4\%\) of the mass energy budget consisting of normal baryonic matter, \(\sim 23–26\%\) consisting of cold dark matter (CDM) weakly interacting with baryonic matter but sensitive to gravity, and the remaining \(\sim 70–73\%\) consisting of smoothly distributed quantum vacuum energy which pushes the Universe apart (Rapetti et al. 2007, Spergel et al. 2007). Several observations (Padmanabhan 2006a, and references therein) indicate that this dark energy component is unclustered and has negative pressure (Rapetti et al. 2007). Its equation of state is \(p = w\rho\), where \(p\) is the pressure and \(\rho\) the energy density. The simplest model for a fluid with negative pressure is the cosmological constant with \(w = -1\) and \(\rho = -p = cSt\); see for instance Peebles & Ratra (2003), Copeland, Sami, and Tsukijawa (2006). So if dark or vacuum energy is a pure cosmological constant \(w = -1\) at all \(z\).

Riess et al. (2004) gave \(w = -1.02^{+0.13}_{-0.19}\) and \(w < -0.76\) at the 95% confidence level, for an assumed static equation of state. Using WMAP (the Wilkinson Microwave Anisotropy Probe) and the SNLS (Supernova Legacy Survey), Spergel et al. (2007) have provided a more recent and precise value of \(w\), assuming that the Universe
is flat ($\Omega_{\text{tot}} = 1$ and $\rho \sim \rho_c$); $w = -0.967^{+0.073}_{-0.072}$ (see also Astier et al. (2006) and Rapetti et al. (2007) who respectively found $w = -1.023 \pm 0.090$ (stat) $\pm 0.054$ (sys) and $w = -1.15^{+0.14}_{-0.18}$). If $w$ is assumed to be $w = -1$ then the deviations from the critical density are small, and the combination of WMAP and SNLS data imply deviations at the level of $-0.011 \pm 0.012$ (Spergel et al. 2007). Even if the prior that the Universe is flat is not included, the combination of WMAP, large-scale structure and SN data can still put strong constraints on the equation of state of dark energy: $w = -1.08 \pm 0.12$ (Spergel et al. 2007). These authors conclude that the three years of WMAP data are consistent with a nearly flat Universe in which the dark energy has an equation of state close to that of the cosmological constant, $w = -1$. They add that the combination of WMAP data with measurements of the Hubble constant, baryon oscillations, supernova data and large-scale structure observations all reinforce the evidence for dark energy.

Generally $w$ might depend on $z$ (non-static equation of state) as discussed for instance by Dicus and Repko (2004), Nesseris and Perivolaropoulos (2007), Jassal, Bagla, and Padmanabhan (2005), Riess et al. (2004), Schaefer (2007), Wang and Dai (2006). Today, the choice between static and dynamic models seems difficult. Riess et al (2004) gave constraints which are consistent with a static nature and inconsistent with a very rapid evolution of dark energy. On the other hand, Jassal, Bagla, and Padmanabhan (2005) noted that whereas WMAP observations allow only a modest variation of energy density of dark energy (DE), the SN data allow a very wide range of variation of the DE density: a factor of 10 between $z = 0$ and $z \sim 1$. Dicus and Repko (2004) concluded that the comparison of models for the equation of state of the dark energy will remain something of “a mug’s game” until substantially more data at higher redshift exist.

Unveiling the nature of dark energy or constraining alternative explanations of accelerated expansion are major issues which demand more data, especially at high redshifts. As GRBs can be observed out to large redshifts and are free from the extinction problem which exists for SNIa (at least for the prompt GRB), they might be complementary candles to SNIa. The Supernova/Acceleration Probe (SNAP—now JDEM, Joint Dark Energy Mission) mission, with its capacity to detect supernovae to $z_{\text{max}} \sim 1.7$, is highly desirable but the use of GRBs which can be easily detected to $z \sim 6$ or more might be a complementary tool in this exciting and promising field of cosmology. There are nevertheless many problems to resolve before GRBs can be used in this way.

The first problem is to check whether GRBs can be considered to be standard candles or at least can be standardized in order to construct a GRB Hubble Diagram (a distance–redshift relation for GRBs). Clearly GRBs are not standard candles. We have seen that their intrinsic isotropic energy spans more than three orders of magnitude and that even the collimation corrected energy spans about two orders of magnitudes, centered around $E_{\gamma} = 6 \times 10^{50}$ erg for a sample of 37 GRBs (see, for example, Figure 2 of Ghirlanda et al. 2004). In fact as we have indicated in Section 3.4, Frail et al. (2001, for 15 GRBs) and Bloom, Frail, and Kulkarni (2003, for 24 bursts) were the first to report a clustering of $E_{\gamma}$ around $5 \times 10^{49}$ erg and $1.3 \times 10^{51}$ erg respectively, suggesting that GRBs are characterized by a standard energy reservoir, despite a very large range of isotropic equivalent energies.
smaller spread found by these authors compared with the Ghirlanda et al. (2004) analysis can be explained by the limited sample of GRBs within a relatively limited range of $E_{\text{peak}}$ ($E_{\text{peak}}$ and $E_\gamma$ being correlated). So, even if one considers the collimation-corrected energy, GRBs cannot be considered to be standard candles. But if correlations can be clearly established there is a real hope for standardizing GRB energies and making them useful as ‘standard candles’ to constrain the cosmological parameters.

Among the first attempts Schaefer (2003) constructed a GRB Hubble Diagram for nine bursts. He used two luminosity (and hence distance) indicators: one proposed by Norris, Marani, and Bonnell (2000) relates the burst luminosity to its spectral lag; the other proposed by Reichart et al. (2001) relates the luminosity to the variability of the burst light-curve. High-luminosity bursts have short lags and spiky light-curves. These properties of the observed light-curves can lead to luminosity estimates and then to the distance of GRBs, a little like the period–luminosity relation of Cepheids.

At the time of his study Schaefer found only nine bursts with the required information: redshift, peak flux, spectral lag, and variability. The Hubble Diagram he obtained was encouraging and promising for GRBs as cosmological tools. Schaefer mentioned the hope of producing with Swift a Hubble Diagram impervious to the effects of dust extinction, out to redshifts that cannot be reached by other methods. As we will see, this enthusiastic and optimistic view of the use of GRBs for cosmography is not unanimously shared. But as we have said the presence of correlations, if they can be firmly established, would be decisive for constraining the Hubble Diagram at high redshift. Such correlations have been found (see Section 3.10). Amati et al. (2002) observed a correlation between $E_{\text{iso}}$ (the energy released in $\gamma$-rays calculated assuming that the emission is isotropic) and $E_{\text{peak}}$, the source frame peak energy, while Ghirlanda et al. (2004) found a correlation between the collimation-corrected GRB energy and the peak energy $E_{\text{peak}}$. Figure 9.1 (from Ghirlanda, Ghisellini, & Lazzati 2004) shows these two correlations. Ghirlanda and co-workers (Ghirlanda, Ghisellini, & Lazzati 2004, Ghirlanda et al. 2004) were very optimistic for future cosmological applications, mainly the measurement of $\Omega_M$ and $\Omega_\Lambda$ in a redshift range not accessible to type Ia SNe: beyond $z = 1.7$. They suggested that GRBs might be the missing link between SNIa and the CMB primary anisotropies. Firmani et al. (2005) noted, however, a lack of low-$z$ GRBs which would be independent of the cosmology. So the Ghirlanda correlation, which is obtained with high-$z$ GRBs, depends on the cosmological parameters that we intend to constrain. This problem of circularity has been discussed by various authors (see, for example, Liang et al. (2008), for an elegant solution to this problem). To ease this problem Firmani et al. (2005) proposed a new Bayesian method that assesses the probability of the observed correlation in various cosmologies. For a flat cosmology they find $\Omega_M = 0.28 \pm 0.03$ for the combined GRB–SNIa data set and the transition redshift between cosmic deceleration and acceleration at $z_T = 0.73 \pm 0.09$. A review of the work of this group is given in Ghirlanda, Ghisellini, and Firmani (2006). Dai, Liang, and Xu (2004) have also used the Ghirlanda correlation for a sample of 12 GRBs to constrain the mass density $\Omega_M$ and the parameter $w$ of an assumed static dark energy equation of state. Subsequently, this group performed a similar analysis of an extended sample of 17 GRBs (Xu, Dai, & Liang 2005).
The critical analysis of these studies by Friedman and Bloom (2005) has to be mentioned. A major concern of these authors is the fact that the quality of the fit to the data with a power law depends strongly on the input assumptions. For instance, the determination of the jet angle \( \theta_j \), which is needed to obtain the collimation-corrected energy \( E_{\gamma} \), is model-dependent. It depends on critical assumptions not only on the density of the circum-burst medium but also on the efficiency of converting the fireball kinetic energy into the radiation emitted during the prompt phase (in both cases the dependence is nevertheless weak). For the uniform jet model in the standard fireball scenario, \( \theta_j \) can be determined by the time \( t_b \) at which the afterglow lightcurve becomes steeper. For a homogeneous circum-burst medium, \( \theta_j \propto t_b^{3/8} \) (Sari, Piran, & Halpern 1999), while for a wind profile, \( \theta_j \propto t_b^{1/4} \) (Chevalier & Li 2000). The main criticisms of Friedman and Bloom (2005) are connected with the assumptions of a constant density medium, of a constant radiative efficiency and of a top-hat jet configuration. Nevertheless, Nava et al. (2006) showed that the Ghirlanda correlation exists in both cases and that it is even tighter in the wind density case. They also showed that it is consistent with the Liang and Zhang relation discussed below. After the results of Swift, however, the achromaticity of the jet breaks has been questioned (see Section 7.5), with the consequence that the connection of \( t_b \) with the jet opening angle is no longer trivial. Friedman and Bloom (2005) also mentioned the unknown

---

\(^2\) For a homogeneous density \( n = \text{cst} \), \( \theta_j \) depends on \( n^{1/8} \) and \( \eta_r^{1/8} \) where \( \eta_r \) is the radiative efficiency. For a wind \( \theta_j \) depend on \( \eta_r^{1/4} \) and on the wind parameter \( A^{1/4} \); \( n(r) = A r^{-2} \) and \( A \), is the value of \( A \) when setting the mass-loss rate due to the wind to \( M_w = 10^{-5} M_\odot \text{yr}^{-1} \) and the wind velocity to \( v_w = 10^3 \text{km s}^{-1} \).
physical basis of the Ghirlanda correlation (but see Rees and Mészáros (2005), showing that dissipative photosphere models can lead to such dependences), the lack of a low-z sample to calibrate the relation in a cosmology-independent way, and several major potential systematic uncertainties and selection effects. All these concerns lead them to urge caution concerning claims of the utility of GRBs for cosmography and attempts to combine the results from GRBs with those of SNe Ia. In any case, and in spite of these objections the authors consider that refinements toward GRB energy standardization might be achieved with Swift. If GRBs do become standardizable candles, tests of cosmological models could be performed up to redshifts $z \sim 10$. Butler et al. (2007) provided more definite conclusions on the possible use of the previous correlations for tests of cosmology. They believed that they are useless and they give three strong, independent reasons to believe that the $E_{\text{peak}}-E_{\text{iso}}$ relation for both Swift and pre-Swift GRBs is an artefact of the detection threshold.

In spite of these pessimistic views let us now discuss two correlations involving three parameters, which have been introduced by Firmani et al. (2006b) and by Liang and Zhang (2005), providing new ways to address the GRB standard candle problem.

Liang and Zhang (2005) considered a sample of 15 GRBs with measurements of their redshift $z$, their spectral energy peak $E_{\text{peak}}$, and the break time $t_b$ of their optical afterglow. They found a strong dependence of $E_{\text{iso}}$ (the isotropic equivalent $\gamma$-ray energy) on $E_{\text{peak}}$ and $t_b$ both measured in the GRB rest-frame. This relationship is purely empirical and, according to Liang and Zhang (2005), it has sufficiently low dispersion to be used as a luminosity indicator. One advantage of this relationship is that it only requires observational data (it does not rely on a model). Liang and Zhang (2005) used this luminosity indicator to measure the distance modulus of GRBs and to constrain the cosmology parameters. They obtained constraints that have large uncertainties due to the small GRB sample. The potential of such three-parameter correlations seems well established, but such tight correlations are intriguing without (presently) understood physical reasons. The authors proposed that $t_b$ might be a unique probe of GRB prompt emission properties. For instance, they suggested that the spectral break in the prompt emission and the temporal break in the optical band may be related to an evolving break in the electron spectral distribution. Hence the break time in the optical band would be different from those in the radio and X-ray bands. As the achromaticity of the jet is not at all proved by the latest Swift results (see Chapter 7), this possibility should be considered. Firmani et al. (2006b) looked for correlations that would not require the determination of $t_{\text{break}}$, which needs follow-up campaigns with large telescopes. They found a three-parameter correlation, which involves only the prompt emission. This correlation involves $L_{\text{iso}}, E_{\text{peak}},$ and $T_{0.45}$; $L_{\text{iso}}$ is the isotropic peak luminosity, $E_{\text{peak}}$ the spectral energy peak, and $T_{0.45}$ is the ‘high signal’ time in the prompt $\gamma$-ray emission. This time, which is the time spanned by the brightest 45% of the total counts above the background, was introduced to measure the variability of GRBs (see Figures 1 and 2 of Reichart et al. 2001 for an illustration). This correlation, found in a sample of 19 GRBs, is given by $L_{\text{iso}} \propto E_{\text{peak}}^{1.62} T_{0.45}^{-0.49}$. The tightness of this correlation encouraged the authors to use it as a new distance indicator for GRBs without redshift. Pseudo-redshifts obtained
for 19 GRBs of their sample are compared to the real redshifts. The histogram of the values of $\Delta z/z$ is compared with the values obtained by Atteia (2003) using another correlation involving $N_e$, the number of GRB photons, $E_{\text{peak}}$, and $T_{90}$, which provided pseudo-redshifts accurate to a factor of 2. Firmani et al. (2006a) used their correlation to standardize the energetics of the 19 GRBs which were combined with 115 SN Ia of the SNLS data set (Astier et al. 2006). To constrain cosmological parameters they used the Bayesian approach described in Firmani et al. (2005) for both populations. They showed that the use of the combined SN Ia+GRB sample reduces the allowed region for the cosmological parameters, compared with the case where a single population is used. Their results indicate that a flat Friedman–Robertson–Walker–Lemaître Λ cosmology (the Λ CDM concordance model) is fully consistent with the Hubble Diagram constructed from the joint sample of SN Ia and GRBs up to $z = 4.5$ (see also Spergel et al. 2007). So Firmani et al. (2007) emphasized the following point: GRBs and SN Ia should not be considered to be competing standard candles but as complementary cosmological probes in two distinct redshift domains. Here again, words of caution have been given by Butler et al. (2007) who found that this correlation could be explained by a bias due to the detection threshold, which prevents the detection of faint hard GRBs that have too few photons.

In spite of this skepticism, the attempts to put constraints on the cosmological parameters have been numerous (see for instance Wang & Dai 2006, Nava et al. 2006, Mosquera Cuesta et al. 2006, Su et al. 2006, Li et al. 2006, Schaefer 2007). Schaefer (2007) evaluated the luminosity distances of 69 GRBs with five luminosity indicators based on correlations involving the burst luminosity: $\tau_{\text{lag}} - L$, $V-L$, $E_{\text{peak}} - L$, $E_{\text{peak}} - E_{\gamma}$, $\tau_{RT} - L$, and $N_p - L$ ($\tau_{\text{lag}}$ is the time lag between the hard and soft light-curves, the variability $V$ quantifies whether the light-curve is smooth or spiky, $E_{\text{peak}}$ is the peak energy of the $\nu F_\nu$ spectrum, $\tau_{RT}$ is the shortest rise time of the GRB light-curve, and $N_p$ is the number of peaks). Schaefer showed that the average of the five luminosity indicators provides a GRB Hubble Diagram which is in agreement with the concordance cosmology (flat Universe with $\Omega_M = 0.27$ and $w = -1$) out to redshift $z > 6$, while SNe Ia are mostly found below $z = 1$ with only 10 of them between $z = 1$ and 1.7 (Figure 7 of Schaefer 2007). This result is quite encouraging. If we follow the conclusions of Schaefer it seems quite certain that GRBs will become a decisive cosmological tool at high redshift ($z > 1$) in the near future with the increasing number of GRBs (see also Mosquera Cuesta et al. 2006) for similar optimistic views). Finally, Schaefer argued that GRBs do not suffer from evolution effects because GRB luminosity indicators are based on light travel times, on energy conservation, and on the degree of relativistic beaming, none of which varies with the metallicity or age of the Universe. So, in spite of the reservations raised by Friedman and Bloom (2005) and by Butler et al. (2007), it seems that the combination of GRBs and SN Ia might become in the future a powerful tool to constrain the cosmological parameters and to study the equation of state and the nature of dark matter.

In the context of these studies, it is worth noting that future GRB missions with cosmological objectives would need to explore a much larger energy range than Swift to be able to determine directly, and precisely, the value of $E_{\text{peak}}$, which is at the center of the various correlations we have reported.
So even though the issue of the role of GRBs as cosmic rulers is not settled, it is highly desirable for the future to keep the determination of cosmological parameters as a major objective of GRB studies, in association with other types of observations like the search for SNe Ia at the highest possible redshifts. Today, however, two fundamental difficulties remain, which have to be surmounted to bring GRBs to the level of other tools.

(1) The scarcity of low-z GRBs does not allow avoiding the circularity problem when standardizing GRBs.

(2) The existence of correlations free of instrument biases is still an open issue. Even if the correlation discovered by Firmani et al. (2006b) is tight and easy to manage because it concerns only the prompt GRB emission, the criticisms which have been raised by Butler et al. (2007) have to be answered and possibly rejected before this correlation can be considered as a robust tool to standardize the GRB energy. This field is still in its infancy and it is important to search for new relations or to confirm the known ones, and like Schaefer (2007), to check the consistency of the known luminosity indicators with an increasing number of GRBs detected by Swift. Finally, it would be crucial to understand the physical origin of whatever correlations are used.

9.2 COSMOLOGICAL PROBES

9.2.1 GRBs and the first stars

At present the GRB redshift distribution extends out to $z = 6.7$ (GRB 080913; see Section 4.10); this section emphasizes the expected impact of highly redshifted GRBs on the knowledge of cosmic star formation back to the epoch of re-ionization, when the first stars were possibly formed. The first stars, called Population III (Pop III) stars, are expected to be quite different from present stars because they are formed from completely metal-free gas, but until now Pop III stars have not been detected. Since GRBs are cosmological events and long GRBs (in this section we will use GRB for long GRBs) are associated with the deaths of massive stars (the collapsar model), they can help elucidate the history of star formation and metal enrichment of the Universe.

As discussed by Loeb and Barkana (2001), in standard cosmological models the early Universe was hot and permeated by a nearly uniform radiation bath. At $z \sim 1200$ the free protons and electrons recombined to form hydrogen atoms, the scattering cross-section dropped precipitously, and most of the photons underwent a last scattering. These photons are observed today as the cosmic microwave background (CMB). The history of the Universe between the recombination of hydrogen at very high $z$ and the formation of the first quasars and stars (at $z \geq 10$) is poorly known. This period has been called the ‘dark ages’ by M. Rees
(see, for instance, Loeb & Barkana (2001) for a review of this period and Lamb & Reichart (2000) for an analysis of the role of GRBs as probes of the high redshift Universe at $z \sim 10$).

The first light from stars and quasars ended the ‘dark ages’ and initiated a ‘renaissance of enlightenment’ in the ‘fading glory’ of the microwave background. At this early epoch the appearance of stars had a dramatic effect on the ionization state and temperature of the rest of the gas in the Universe. In fact it seems that the conversion of a small fraction (i.e. $10^{-5}$) of the total baryonic mass into stars or black holes could have been sufficient to completely ionize the rest of the Universe (Loeb & Barkana 2001). The redshift of this epoch of re-ionization is not well defined because calculations of structure formation in the popular CDM (cold dark matter) cosmologies imply that the Universe was ionized at $z \sim 6–12$ (Ostriker & Gnedin 1996, Gnedin & Ostriker 1997, Haiman & Loeb 1998, Valageas & Silk 1999, Chiu & Ostriker 2000, Loeb & Barkana 2001). Gnedin and Ostriker (1997), for instance, gave $z \sim 7$ assuming that the re-ionization is caused by the first generation of Pop III stars. Haiman and Loeb (1998), on the other hand, predicted a re-ionization occurring at $z > 11.5$, if it is caused by an early population of low-luminosity quasars and/or stars. This re-ionization was preceded by a period of reheating between $z = 20$ and $z = 7–10$, both periods being due to a high-energy radiation field. Reheating is a rather slow process, followed by sudden re-ionization (Gnedin & Ostriker 1997). Future observations are needed to identify when and how the intergalactic medium was ionized. GRBs might be valuable tools to illuminate this period, due to the fact that quasars are becoming much rarer at these high redshifts, while GRBs might have started earlier (and they have already been detected beyond $z = 6$). GRB 050904 at $z = 6.295$ was probably the first GRB to be possibly identified at the epoch of re-ionization. Hence GRBs at high $z$ might provide a unique opportunity to investigate star-forming regions at these early times (Gou, Fox, & Mészáros 2007), and possibly beyond, when the first generations of stars started to form, perhaps around $z = 30$ (Abel, Bryan, & Norman 2002). It is worth noting that in spite of its great distance GRB 050904 had bright prompt gamma-ray emission and early infrared and radio afterglows. Moreover such events are not too rare since three bursts at least have been detected with $z > 5$ in the first 2 years of Swift observations: GRB 050814 at $z = 5.3$ (Jakobsson et al. 2006), GRB 060522 at $z = 5.11$ (Cenko et al. 2006), and GRB 060927 at $z = 5.6$ (Fynbo et al. 2006a). Another example was the detection of GRB 080913 at $z = 6.7$. Although the fraction of GRBs expected at high redshift is model-dependent, it is clear that it can be significant. For instance Natarajan et al. (2005) found that 25% to 40% of GRBs might appear at $z > 4$. For the higher value the authors relate the GRB rate to a simple model of the average metallicity as function of $z$ and they assume that GRBs appear preferentially in low-metallicity environments. Bromm and Loeb (2006) considered that a robust value for the fraction of GRBs at $z > 5$ is 10% of all Swift bursts (Figure 9.2). Jakobsson et al. (2006) evaluated this fraction to be between 7% and 40%, and Le and Dermer (2007) gave 8% to 12%. Gou et al. (2004) indicated that Swift could detect GRBs up to $z \sim 30$ if they are present. Daigne, Rossi, and Mochkovitch (2006) predicted the redshift distribution of long GRBs using Monte Carlo simulations taking into account
Figure 9.2. Predicted GRB rate (GRBs per year) for Swift as a function of redshift (from Bromm & Loeb 2006). (a) Late re-ionization ($z \sim 7$). Dotted lines show the contributions of Pop I/II and Pop III stars to the observed GRB rate, for the case of weak chemical feedback. Dashed lines show the contributions of Pop I/II and Pop III stars to the observed GRB rate, for the case of strong chemical feedback. Newly created metals are dispersed into the IGM via SN driven winds. The cases of weak and strong chemical feedback correspond to winds experiencing large and small radiative losses respectively (Bromm & Loeb 2006). (b) Early re-ionization ($z \sim 17$). The different lines have the same meaning as in panel (a). Filled circle gives the GRB rate from Pop III stars if these were responsible for re-ionization of the Universe at $z = 17$. For the two cases the GRB rate from Pop III progenitors appears very uncertain in comparison with the rates from Pop I/II.
instrumental limitations. They reported an evolution of the efficiency of GRB production by massive stars that would be six to seven times higher at $z \sim 7$ than at $z \sim 2$. Thus GRBs would not directly trace the star-formation rate (SFR) (see also Le Floc’h et al. 2006). They predict the detection of $\sim 10$ GRBs at redshift $z > 6$ for a 3-year Swift mission, based on a yearly rate of $\sim 100$ Swift GRBs. This is somewhat less than previous values, but there is substantial uncertainty in this number. In any case as the authors noted, the fraction of high redshift GRBs is significant and might be sufficient to open a new observational window on the high-redshift Universe.

GRBs might therefore be used to study the first generation of stars. Abel, Bryan, and Norman (2002) indicated that while the mass of the first stars remains unclear, a realistic minimum mass should be $> 30 M_\odot$. Their simulations would imply the possibility that all metal-free stars were massive and formed in isolation. The SNe associated with these first massive stars would release the heavy elements which were absent in the first generation of stars. As the stellar evolutionary timescale for these massive stars is short ($\sim 10^6$ yr) the enrichment of the IGM with heavy elements could have occurred rather promptly (Bromm & Loeb 2006, and references therein). Thus, substantial metal enrichment is expected at high redshift in star-forming regions with $Z/Z_\odot \sim 1/30$ in the high-density regions, while essentially no metals were present in low-density regions (Gnedin & Ostriker 1997). This would explain the absence of completely metal-free low-mass stars in the Milky Way. These earliest generations of stars would occur at redshift $z \sim 16$–20, with star formation peaking at $z \sim 2$–10 (Gnedin & Ostriker 1997, Valageas & Silk 1999). These massive Population III stars with zero or very low metallicity would be very good candidates for GRBs (see the conditions for forming a collapsar in Chapter 8). These possible high-redshift GRBs ($z > 7$, beyond the current horizon of galaxy and quasar surveys) would trace the first stars illuminating this otherwise dark Universe (Abel, Bryan, & Norman 2002).

However, Pop III stars can be viable progenitors of long GRBs triggered by the collapsar mechanism only if they lose their outer envelope, for instance through mass transfer to a companion star in a close binary (Bromm & Loeb 2006). But Langer and Norman (2006) considered that the low metallicity bias of GRBs reduces the probability that binary evolution channels can be significant GRB producers. Woosley and Heger (2006) studied massive, rapidly rotating, single stars as progenitors of GRBs. The low metallicity would imply fast rotation, so long GRBs would prefer low-metallicity environments (with the majority of GRBs formed in regions with a fraction of solar metallicity) and single massive stars might be possible GRB progenitors (see also single-star models with high mixing proposed by Yoon and Langer 2005 and WO stars only produced at low metallicity proposed by Hirschi et al. 2005). This dependence on metallicity would result in GRB formation peaking at significantly higher redshift than supernovae (Langer & Norman 2006). This single massive star possibility is in contradiction with the conclusions of Bromm and Loeb (2006) who consider that isolated massive Pop III stars are most likely not able to shed much mass prior to their final collapse and cannot give rise to a collapsar-driven GRB. This is the reason why they consider binary systems going through a common envelope
phase as better candidates. Another problem is the possible number of GRB progenitors. Berezinsky et al. (2007) found that there are not enough fast rotating stars in Pop III binaries; they predicted a detection rate which is quite low, \( \sim 0.1-0.01 \, \text{yr}^{-1} \). If we take into account all these objections, we can conclude that the production of GRBs from Pop III stars (isolated or in binary systems) is quite an open issue, which clearly demands better models of zero metallicity Pop III stars.

### 9.2.2 GRBs and quasars

As we have seen, GRBs might represent a good tool to understand the first generation of massive stars. They would provide constraints on the initial stellar mass function and on the chemical enrichment history of galaxies hosting GRBs. Until now the chemical enrichment of the Universe was studied with the observation of strong absorption lines detected in bright high-z quasars (Wolfe, Gawiser, & Prochaska 2005). Such absorption systems, called damped Lyman \( \alpha \) (DLA) systems, are due to intervening galaxies along the quasar (QSO) line of sight (QSO–DLAs). Long GRBs are generally followed by optical and X-ray afterglows which can briefly exceed the absolute brightness of any known quasar by orders of magnitude, if they are observed very early after the GRB. Ten minutes after a GRB the optical afterglow can be as bright as magnitude 14–19, decreasing a few hours later to magnitude 16–22 (Kann et al. 2007). These afterglows can serve as bright background sources for probing intervening gas along the line of sight.

Let us consider the absorption lines of the afterglows. They provide the column density of elements, i.e. the number of atoms per unit surface which are responsible for the absorption. The line width is related to the temperature and to the level of turbulence of the absorbing gas. As the GRB line of sight usually crosses complex regions, multiple absorptions are observed, and fine spectroscopy is highly desirable to resolve them. From the observation of UV absorption lines redshifted into the optical region (this is the case for the majority of GRBs with a measured redshift) the column densities of different elements are obtained after a dust depletion correction has been applied to take into account the elements locked onto dust grains, which escape UV detection. On the other hand, X-ray observations of the afterglow might provide the total column density because the absorption in X-rays is not sensitive to the element state (see, for example, Watson et al. (2006, 2007), Guidorzi et al. (2007), and Section 9.2.4).

We will concentrate first on the spectroscopy of the afterglow with its absorption lines as a tool to provide a measure of the metallicity, dust content and kinematics of the interstellar medium at the location of the GRB (most often a star-forming region in the host galaxy). The spectroscopy of GRB afterglows may also provide direct information on the nature of the massive progenitors themselves (Berger et al. 2006). In contrast to what is generally possible with quasars, the host galaxy can be observed when the bright emission of the GRB is gone a few days after the burst. Thus, late-time deep imaging of the GRB host galaxy and intervening galaxies along the sight lines is possible without the contribution of a bright background quasar, as in QSO–DLA systems. As long GRBs appear to be formed in active star-forming
regions, the GRB host galaxy with its emission lines can be detected even if it is faint and small because UV photons coming from intense star formation excite the ionized gas (H II region). It is therefore possible to conduct efficient studies of the ionization state and metallicity of the general ISM in GRB host galaxies.

Hence through the spectroscopy of their early afterglows and at a later time of their host galaxies, GRBs are as powerful and independent tools to complete the study of QSO–DLAs. They are specifically adapted to the study of the regions in which intermediate to high redshift star formation occurs and in particular to the studies of the physical, chemical, and dynamical state of the ISM in star-forming GRB galaxies, with several advantages over quasar studies (Penprase et al. (2006):

- First, as GRBs are embedded in the regions of most intense star formation, they are directly linked to star formation and they provide clues to the production and dispersal of metals.
- Second, since the GRB afterglow emission fades away on a timescale of days to weeks the host galaxy can be studied directly.
- Third, as GRBs are located within star-forming regions, they provide the only systematic way to directly probe the small-scale environment and conditions of star formation at high redshift, whereas the probability of intersecting an individual star forming cloud in a quasar line of sight is vanishingly small.

Of course it would be fruitful to perform these various types of investigations (absorption line and emission line searches) for the same burst; unfortunately, until 2007 this had not been the case. Moreover, due to the limited afterglow duration and the need to observe it when it is bright (to do high-resolution spectroscopy) it is mandatory to observe as soon as possible after the GRB trigger, using large telescopes with high-resolution spectrographs with a wide spectral range. Finally, it is clear that NIR spectroscopy is needed to study high-redshift GRBs.

The interest of GRBs for probing the environment of the GRB site, the ISM of star-forming host galaxies, and the intergalactic medium has led to numerous studies. In the following sections, we briefly discuss the metallicity of the host galaxies and of intervening galaxies. We will refer to the review of Savaglio (2006) in the New Journal of Physics describing the knowledge of the chemical enrichment in GRB environments as of 2006 (this is a rapidly evolving field). Emphasis is placed on the metallicity measured through rest-frame UV absorption lines and associated with the neutral cold gas ($T < 10^3$ K) distributed along the GRB line of sight. This is called the metallicity of the GRB–DLA. Comparison is made with the results derived using QSO–DLA studies. Savaglio (2006) also discussed the metallicity measured using emission lines originating in the ionized warm gas ($T \sim 10^4$ K) of star-forming regions in the entire host galaxy. This is the host galaxy metallicity. We will just emphasize the main conclusions of the author, for these two rich fields of investigations. For a complete analysis of this new field of GRB studies and for a detailed list of the different contributions the reading of Savaglio’s review is illuminating.
9.2.3 Afterglow spectroscopy: absorption lines of the intervening matter

Since most Swift GRBs have a redshift \( z > 0.5 \), the UV absorption lines are redshifted into the optical domain (the low redshifts will remain out of reach until the installation of the next COS UV spectrograph on HST in 2009). The relative abundances of the different elements, which can be deduced from the study of the optical afterglow, permit gathering data on the chemical enrichment and dust content of the gas along the GRB line of sight and thus to study nucleosynthesis in the GRB star-forming environment. In her Table 1, Savaglio (2006) reported \( \text{H} \) and heavy element column densities measured in 22 GRB afterglows, i.e. more than one-quarter of all GRBs with measured redshift. Whereas it is clear that the statistical analysis of physical parameters and their evolution with \( z \) is premature (the sample is still too limited), interesting trends can nevertheless be mentioned. For instance, the comparison with QSO–DLAs indicates larger column densities of hydrogen and heavy elements for GRB–DLAs (Figure 9.3). The majority of GRB hosts have large neutral gas content with neutral hydrogen column density beyond \( \log N (\text{H} \text{I}) = 20.3 \), the standard threshold defining DLAs. The metallicity (heavy elements relative to hydrogen) is higher than in the cold ISM of normal high redshift galaxies. Hence, in spite of a small sample, it is tempting to conclude that GRBs might be the only objects allowing the detection of very dense star-forming regions at high redshift.

Dust properties can also be studied (Savaglio 2006). As we have pointed out UV absorption lines account for the gas phase only, so measuring heavy element abundances requires a dust correction. Dust content in a GRB–DLA can be characterized by comparing abundances of elements with different depletion characteristics, iron-to-zinc relative abundance being the best diagnostic. Figure 3 of Savaglio (2006) compares [Fe/Zn] and [Si/Zn] for QSO–DLAs and five GRB–DLAs as a function of Zn II column density. In this figure GRB–DLAs and QSO–DLAs are located in two different regions suggesting that they belong to two distinct populations, but the GRB sample is really too small to be confident of this. If, as expected, (Fe/Zn) measures the dust depletion, this figure indicates that iron in GRB–DLAs would be largely locked into dust grains, up to 90–99% (Savaglio 2006).

![Figure 9.3](image)

**Figure 9.3.** Fraction of GRB–DLA (filled histograms) and QSO–DLAs (empty histograms) per HI, FeII, and Zn II column–density interval (left, middle, right panels respectively). These histograms show that column densities in GRB–DLAs are on average higher than in QSO–DLAs (Savaglio 2006).
Figure 9.4. Depletion pattern in five GRB–DLAs including four or more heavy elements (from Savaglio 2006). For GRB 000926 and GRB 050401 the Mn column density is not available, dotted lines suggest its possible values. The dashed line at the bottom is the mean depletion pattern measured in 20 QSO–DLAs for which all six elements are detected. This figure shows that metal column densities and dust depletion in GRB–DLAs are much higher than in QSO–DLAs. Zn-to-Fe relative abundance, which represents dust depletion, is on average four times larger in GRB–DLAs than in QSO–DLAs. The inset gives the dust-depletion patterns of the Galactic ISM (from top to bottom are indicated different situations taken from Savage and Sembach (1996): no dust, warm halo, warm halo + disk, warm disk, cool disk clouds).

2006). Let us discuss now the dust-depletion pattern and the dust correction in more detail.

Dust-depletion pattern

As different elements are locked in dust grains by different amounts, the column densities carry the signature of the chemical composition of the dust, i.e. the dust-depletion pattern. In the ISM of the Milky Way the depletion pattern depends on the gas properties, such as its density and temperature. The dust depletion is larger for denser and cooler clouds (see Savage & Sembach (1996), and the inset of Figure 9.4 showing the dust depletion patterns of the Galactic ISM). The results obtained for the five GRB–DLAs resemble the depletion patterns observed in the Milky Way for a warm disk + halo or for warm disk gas, supporting the presence of dust in GRB–DLAs. The comparison with QSO–DLAs is also given in Figure 9.4. For instance the relative abundance of Zn/Fe (characteristic of dust depletion) is on the average four times larger in GRB–DLAs than in QSO–DLAs.

Dust extinction

Dust extinction can be due to intervening material in the vicinity of the GRB or in the host galaxy, possibly across several kiloparsecs, and sometimes also due to interven-
Figure 9.5. Distribution of the host galaxy visual extinction $A_V$ in the source frame for the Golden Sample of 19 GRB afterglows. The data point at $A_V = 3.8$ is the lower limit to the visual extinction derived for GRB 970828. (Djorgovski et al. 2001). There are no afterglows with $A_V > 0.8$, creating a dark burst desert. The inset gives a zoom into the $A_V < 0.8$ region. The distribution peaks near $A_V = 0.15$, showing that even if it is small there is a definite amount of extinction along the line of sight to GRB afterglows. This indicates that often afterglows are not obscured by dust close to the burster (from Kann, Klose, & Zeh 2006).

ing material between the host galaxy and the observer. Dust extinction is usually called reddening because UV radiation is more affected than optical radiation; it is often expressed in terms of extinction in the visual band $A_V$. The extinction is proportional to the column of dust along the line of sight. The dust extinction law $A_V$ is well known only for the Milky Way (MW) and for the Large and Small Magellanic Clouds (LMC, SMC). Extinction curves can be derived by fitting the observed optical/NIR spectral energy distributions (SEDs) of GRB afterglows with a power-law spectrum, reddened by an extinction law of known type, mostly MW, LMC, or SMC. Among these three dust extinction models (MW, LMC, SMC), the prevalence of SMC-like dust in GRB host galaxies was observed by Kann, Klose, and Zeh (2006) for the Golden Sample of 19 GRB afterglows, and by other groups for various bursts (see references in Kann 2006), although MW dust models cannot be excluded. Kann, Klose, and Zeh (2006) derived the host galaxy extinction $A_V$ along the line of sight and the intrinsic (extinction corrected) spectral slope $\beta (F_\nu \propto \nu^{-\beta})$ of the afterglow in optical and NIR bands. The distribution of $A_V$ is given in Figure 9.5 (Kann, Klose, and Zeh 2006). No afterglow with $A_V > 0.8$ is observed, creating a dark burst desert (Kann, Klose, and Zeh 2006). The inset of Figure 9.5 shows that a small but definite amount of extinction is present along the line of sight. These results seem to indicate that afterglows are usually not obscured by dust close to the bursters. Figure 9.6 (Kann, Klose, and Zeh 2006) gives the dust-to-gas ratio ($A_V$ as a function
Figure 9.6. Dust-to-gas ratios in the host galaxies along the line of sight to the GRBs of the Golden Sample; $N_H$ values are taken from Kann, Klose, and Zeh (2006, and references therein). The $N_H - A_V$ relations are from Bohlin, Savage, and Drake (1978), Predichl and Schmitt (1995), and Hopkins et al. (2004). Most lines of sight to the GRBs have a dust-to-gas ratio that is even lower than the SMC. Two exceptions in this Golden Sample are GRB 021004 and GRB 030329, with high dust-to-gas ratio. The $N_H$ values for these two bursts are only upper limits (Zeh, Klose, & Kann 2006). Figure from Kann, Klose, and Zeh (2006).

of $N_H$) in the host galaxies. Most GRB lines of sight have a dust-to-gas ratio which is even lower than the SMC.

Savaglio (2006) estimated the visual extinction $A_V$ by comparing the amount of metals distributed along the GRB line of sight with the same quantities observed in the local Universe for which $A_V$ is measured. As a reference point she considered $A_V$ measured in the ISM of the LMC and she proposed a method to estimate $A_V$ when Fe II and Zn II or only Zn II are available. This method is considered by the author as a simplification of the method used by Savaglio and Fall (2004) to obtain $A_V$ for GRB 020813. For this GRB, they studied simultaneously for the first time the dust depletion pattern and the extinction law in the circum-burst medium. This was possible thanks to an excellent optical spectrum (Figure 9.7), in general very difficult to obtain. The column density of six heavy elements was measured. The relative abundances resemble the Galactic dust depletion pattern from which a visual extinction $A_V \sim 0.40$ was obtained. This high visual extinction was difficult to reconcile with the observed UV GRB continuum spectrum which deviates very little from a perfect power law, suggesting low reddening. A way of avoiding this difficulty was to assume an extinction curve which is weakly dependent on the wavelength (grey extinction). But we will see later that other explanations have been proposed.
Figure 9.7. Optical spectrum (λλ = 1420–4080 Å rest-frame) of the afterglow of GRB 020813 at \( z = 1.255 \), after correction for galactic extinction with \( E_{B-V} = 0.111 \). The straight line is the power-law fit. At that time it was one of the best spectra in terms of resolution and signal-to-noise ratio; it was obtained with the LRIS spectrograph on the Keck I telescope, \( 4.65 \) h after the HETE-2 detection, while the visual magnitude was \( \sim 19 \) (from Savaglio & Fall 2004).

Savaglio (2006) noted that it would be reasonable to expect high dust extinction, given the large metal column densities generally measured in GRB–DLAs, but surprisingly, GRB afterglows often show a featureless power-law behavior in the rest-frame UV domain (Chen et al. 2005, Barth et al. 2003). This lack of reddening disagrees with the possible presence of high dust depletion (see also Savaglio, Fall, & Fiore 2003 for similar conclusions). But this can be explained if the dust is distributed in clumps, because in this case the GRB afterglow optical emission may be weakly affected by the presence of dust (Witt & Gordon 2000).

To summarize, in GRB afterglows \( A_V/N_H \), the ratio of extinction and/or reddening to neutral H (HI) column densities (determined from X-ray or Ly\( \alpha \) absorption) is smaller than that of the Milky Way by a large factor (Galama & Wijers 2001, Vreeswijk et al. 2004, Hjorth et al. 2003). This effect, reported for the first time by Galama and Wijers (2001), has been explained by dust destruction out to a distance of \( \sim 10–20 \) pc (see also Waxman & Draine 2000, Fruchter, Krolik, & Rhoads 2001, Draine & Hao 2002, Perna & Lazzati 2002, Perna, Lazzati, & Fiore 2003). The modification of the extinction law would be due to the preferential destruction of small grains by UV and X-ray radiation from the GRB (giving a grey extinction law; see also Stratta et al. 2005, Savaglio & Fall 2004). It might also be the result of the growth of dust through coagulation in high-density environments, such as the cores of star-forming regions harboring GRBs. Both effects, by the destruction of small grains, lead to a dust size distribution that produces a flat extinction curve weakly dependent on wavelength. Such flat extinction curves imply less reddening (Chen, Li, & Wei 2006, and references therein, Waxman & Draine 2000, Fruchter, Krolik, & Rhoads 2001, Perna, Lazzati, & Fiore 2003).
For GRB 020124 at $z = 3.20$ Hjorth et al. (2003) suggested another possibility: GRBs could occur in less chemically enriched environments, similar to those of the SMC and QSO–DLAs. The low value of the ratio between the optical extinction ($A_V < 0.2$) and the column density of neutral hydrogen suggests that the GRB environment is low in dust because of its low metallicity (e.g. for a universal dust-to-metal ratio) and/or because of a low dust-to-metal ratio, a large part of the metals being in the gas phase. The dust-to-metal ratio measured in GRB afterglows is less than 11% of that found in the MW, but consistent with the SMC and high redshift QSO–DLAs. Pettini et al. (1997a,b) arrived at the same conclusion for QSO–DLAs, which are generally metal-poor at all redshifts sampled and have a typical low optical extinction, $A_V \leq 0.1$, for most of them. Watson et al. (2006) also indicated that a non-universal dust-to-metal ratio would explain a high metal column density with little spectral reddening without requiring non-standard extinction laws, since there would simply be less dust, as proposed by Hjorth et al. (2003). Watson et al. (2006) found that a low dust-to-metal ratio is a very natural possibility for GRB hosts given their high $z$ and very young stellar populations. This is because metal enrichment of the ISM depends on SNe (which have short lifetimes) while dust is not produced by SNe but mainly by asymptotic giant branch (AGB) stars with lifetimes at least 100 Myr (Andersen, Höffner, & Gautschy-Loidl 2003), i.e. there would be a long delay before the most massive stars begin to produce large amounts of dust (Ferrarotti & Gail 2005).

To summarize, we can say that the lack of reddening can be due to a flat extinction curve and a lot of dust or to a lack of dust and a reasonably steep extinction curve. These two possible explanations might be distinguished by X-ray spectroscopy of high-$z$ GRBs combined with accurate optical/NIR observations of the afterglow (Hjorth et al. 2003). Moreover, as noted by Savaglio (2006), correcting the GRB observed flux using the extinction law known for the local Universe, though it is today the best that can be done, is only a rough approximation that needs to be tested by other means. The critical missing element is an extinction curve derived from GRB afterglows, which will avoid these permanent references to the nearby environment (i.e., the Milky Way, LMC, and SMC). This has been done by Chen, Li, and Wei (2006), Li et al. (2008), and Liang and Li (2008), for instance. These authors derived the extinction curve of various GRB host galaxies without any a priori assumption of a specific extinction law. This approach is expected to provide a better knowledge of the dust and extinction properties of GRB host galaxies which are still poorly known.

Thus the question of dust and extinction in GRB environments is still an open issue.

**Metallicity**

With their golden sample of 19 afterglows covering the redshift range $z = 0.1–4$, Kann, Klose, and Zeh (2006) found that the majority of GRBs occur in low-metallicity environments (see also Vreeswijk et al. 2004, Chen et al. 2005, Chen, Prochaska, & Bloom 2006, Jakobsson et al. 2004, Starling et al. 2005, Prochaska et al. 2004).
Figure 9.8. Redshift evolution of metallicity for nine GRB–DLAs (●) and 197 QSO–DLAs (□). The solid and dashed lines indicate the best-fit linear correlation for GRB–DLAs and QSO–DLAs respectively. GRB–DLA metallicity is on average about five times larger than QSO–DLA metallicity. The age of the Universe is indicated on the upper horizontal axis (from Savaglio 2006).

After a dust-depletion correction that takes into account the fact that some of the heavy elements are partially locked into dust grains, the metallicities of nine GRB–DLA have been calculated by Savaglio (2006). She derived the evolution of metallicity versus redshift for these GRB–DLAs and 197 QSO–DLAs (Figure 9.8). Despite the small size of the sample and the large dispersion (from 0.02 solar to near-solar metallicity), it is clear that GRBs have a higher metallicity than most QSOs, and it might be that the evolution of GRB–DLA metallicity with redshift is flatter than that of QSOs (see also Berger et al. (2006) and Fynbo et al. (2006b) for similar studies).

The question of GRB metallicity is largely debated because Hirschi, Meynet, and Maeder (2005) and Woosley and Heger (2006) found that massive single stars may become GRB progenitors only if their metallicity is below \( \sim 0.3 \, Z_\odot \) (low metallicities decrease the mass-loss and make it possible to retain a high angular momentum). So knowing the metallicity of the GRB environment is a crucial objective to check this point. Of course, if GRBs are found in \( Z > 0.3 \, Z_\odot \) environments, the collapsar model is not dead, because binary systems can lead to collapsar explosions without conditions on the metallicity of the progenitor (see Chapter 8). Figure 9.8 shows that indeed most metallicities from GRB absorption systems are below 0.3 \( Z_\odot \). Nevertheless, this figure also shows GRBs above \( Z = 0.3 \, Z_\odot \); for example, GRB 000926 (\( Z = 0.7 \, Z_\odot \); Savaglio, Fall, & Fiore 2003), and GRB 980425 (\( Z = 0.8 \, Z_\odot \); Sollerman et al. 2005, not shown in the figure). These exceptions indicate that single massive stars might produce GRBs even if \( Z > 0.3 \, Z_\odot \) or, if 0.3 \( Z_\odot \) is really a hard limit for massive single stars, that some GRBs have binary progenitors (but see the critical remarks of Prochaska (2006) on metallicity determinations, at the end of this paragraph). For the future it will be necessary to confirm that GRB–DLAs and
QSO–DLAs really represent two different classes of absorbers though they both trace the ISM in high-z galaxies. This difference might be explained if, as suggested by Savaglio (2006), one takes into account the fact that GRB–DLAs are less affected by observational limitations and that GRBs are produced in regions of active star formation with metal enrichment that could be faster than in average galaxies at the same redshift. As we have already mentioned, the difference could also be explained because QSO absorption lines involve galactic halos along the line of sight probing mainly low-density regions while GRB–DLAs probe denser regions of the ISM, active star-forming regions and molecular clouds, so their metallicity is naturally higher. It has also been suggested that QSO–DLAs would preferentially probe low-mass galaxies (Savaglio et al. 2005), which are less metal-rich (Lee et al. 2006).

Before drawing any conclusion on this topic a word of caution about afterglow spectroscopy is needed. As the majority of observations to date were obtained with low or medium resolution spectrometers, Prochaska (2006) argued that the multi-ion single-component curve-of-growth analysis technique, which is generally used, under-estimates the column densities of metal–line profiles observed in GRBs. He concluded that the majority of abundances reported in the literature are underestimated with reported errors frequently over-optimistic. This analysis implies that profound consequences have to be expected in the determination of the metallicity, dust-to-gas ratio and chemical abundances of the ISM in GRB host galaxies. As an example, Prochaska (2006) indicated the minimum and maximum values of the metallicity allowed by the data for 10 GRBs with redshift greater than 1.255. Solar and even super-solar abundances are allowed in nearly every case. It is also interesting to note that he considered that the common conclusion that GRBs occur in metal-poor ISM may need serious revision. Savaglio (2006) also analyzed the possible systematic effects and particularly the curve-of-growth technique. She recognized that this issue is a matter of controversy but she noted that if the GRB–DLA metallicity is underestimated this increases the difference between GRB–DLAs and QSO–DLAs. Savaglio (2006) discussed other systematic effects which could compromise the reported conclusions on the metallicity derived for GRB-DLAs. In conclusion, the question of metallicity and dust content along the line of site of GRBs is not trivial and needs additional work to be fully understood.

Until now we have restricted the studies of absorption lines to the optical/NIR range, but the X-ray spectra of GRB afterglows also contain the signature of absorption. These X-ray spectra can be used to determine metal column densities, which can be compared with values measured with optical/NIR afterglows.

### 9.2.4 Soft X-ray absorption

Soft X-ray absorption provides a direct measure of the total column density of metals in the solid and gas phases. Hence the problem of dust is avoided. X-ray absorption is dominated by $\alpha$-chain elements. These $\alpha$-elements have atomic numbers that are multiples of the atomic number of the $\alpha$-particle He, such as O, Si, S, Ca, and Mg. The biggest contribution to the absorption comes from the oxygen K shell absorption
edges with OI at 0.52 keV (Watson et al. 2006). The use of these α-elements in GRB–DLAs (hence in the optical domain) has been discussed by Savaglio (2006): S and Si are observed, and sulfur (S) is an interesting element for estimating the metallicity because in the galactic ISM it is not affected by dust depletion, but due to the scarcity of the observations no definite conclusion on its abundance can be drawn.

X-ray measurements do not suffer from the critical problems which plague optical observations: the need for high-resolution spectra and the assumptions regarding dust depletion. However, as noted by Watson et al. (2007), combining X-ray metal column densities with hydrogen column densities is very difficult due to the conflicting redshift requirements: optical observations require high redshift ($z > 2$) GRBs for the detection of the hydrogen Lyα line (1216 Å) which is otherwise in the UV, while low redshifts are needed in order to have the soft X-ray absorption visible within the bandpass of X-ray instruments. An example of this use of optical and X-ray afterglows is given in Watson et al. (2006, 2007).

Galama and Wijers (2001) were the first to raise the $N_H - A_V$ problem, using X-ray data from BeppoSAX. They found that metal column densities measured in X-rays were too large in comparison with the values expected from the observed reddening, assuming Galactic, LMC or SMC metal-to-dust ratios. Watson et al. (2006) considered GRB 050401 and combined the optical afterglow spectrum and the early X-ray spectrum from the Swift XRT. This was the first direct measurement of the H I column density and of the total metal (primarily α-element) column density in a single GRB. They computed $A_V$ from the optical to X-ray spectral energy distributions (SEDs), independent of the shape of the extinction curve: they obtained $A_V < 4.5$, well below the dust column that is inferred from the X-ray column density: $A_V \sim 9$. This discrepancy between a high metallic absorption and a low total extinction might be explained by small dust content with high metallicity (low dust-to-metal ratio) but it cannot be explained by grey extinction (Watson et al. 2006). Little dust with high metallicity may be produced by sublimation of dust grains. This is puzzling, however, because at the same time there is some evidence of dust-depletion effects in the ratios of the optical lines (Watson et al. 2006).

Using optical spectroscopy at the William Herschel Telescope and Swift XRT observations to look for evidence of intrinsic absorption in the X-ray spectrum, Starling et al. (2005) found that the host galaxy of GRB 050730, at redshift $z = 3.968$, is a low-metallicity galaxy with low dust content. The X-ray absorbing gas is situated close to the GRB (see Watson et al. 2007) while the HI absorbing causing the DLA is most likely located farther out, unaffected by GRB radiation (Starling et al. 2005). Watson et al. (2007) revisited the comparison of H I column densities from Lyα absorption (optical afterglow) with the metal-column densities from soft X-ray absorption (X-ray afterglow). They considered eight bursts. The absorbing column densities obtained by the two methods, using UV and X-rays, are compared in Figure 9.9. This figure clearly shows that there is no correlation between optical and X-ray column densities. Moreover, the optical to X-ray ratio spans a range of at least several hundred and the metal abundances derived from this comparison indicate metallicities far above the solar values (Figure 9.9). These
Figure 9.9. Neutral hydrogen column density (measured with hydrogen Lyman $\alpha$ absorption) as a function of the X-ray equivalent hydrogen column density (assuming standard metallicity). The ratio of these values provides an estimate of the metal abundance of the absorbing medium. The lack of correlation between these two column densities seems to indicate strong variation in the metal enrichment of the absorbing clouds near GRBs. The values are so extreme, with many which are much greater than $Z_{\odot}$, that they led Watson et al. (2007) to consider the possibility of strong systematic uncertainties related to either or both of these absorbing column density estimates. These uncertainties could be related to the ionization of most of the hydrogen along the line of sight. The Crab pulsar and the blazar AO 0235+164 are shown for comparison (from Watson et al. 2007).

two results led the authors to conclude that the X-ray absorption and the Ly$\alpha$ absorption are sampling different environments. After eliminating various possibilities, Watson et al. (2007) proposed solving this issue by considering the ionization of the gas cloud surrounding the GRB by the early UV afterglow out to great distances along the line of sight. The X-ray absorbing metals, while also ionized, will not cease to absorb soft X-rays substantially. This possibility may help to resolve the $N_H-A_V$ problem. With this concept of ionization it can be expected that optical/UV data are affected by non-ionized material at great distances (with a dust-to-gas ratio more akin those found in high-$z$ DLAs), whereas the X-rays probe ionized material closer to the GRB. Watson et al. (2007) estimated the size of the ionized cloud to be less than $\sim$3 pc. UV/optical spectra are probing gas that is largely outside this bubble, in which the GRB explodes. If this conclusion is confirmed, it means that the optical afterglow probes the real ISM in the host galaxy, and therefore is not
strongly affected by the properties of the circum-burst region and by the impact of the progenitor on its surroundings.

### 9.2.5 Fine-structure transitions

Another feature of GRB environments is the presence of strong fine structure transitions from excited states of C\(^+\), Si\(^+\), O\(^0\), and Fe\(^+\). Such absorption lines are useful to study the physical state of the gas (Bahcall & Wolf 1968). They can reveal the temperature and density of the gas. For instance, identifications of strong Fe\(\text{II}\) fine structure transitions suggest extreme gas densities and temperatures in the GRB environment, similar to what is found around massive stars like luminous blue variables (LBV) and Wolf–Rayet stars (Chen, Prochaska, & Bloom 2006).

Vreeswijk et al. (2004) using the fine structure lines Si\(\text{II}\) at \(\lambda = 1309\) and 1533 Å estimated the H\(\text{I}\) volume density of the gas producing this absorption, \(n_{\text{HI}} = 10^2 - 10^4\) cm\(^{-3}\), under the assumption that these fine structure lines are populated by collisions. This volume density is higher than that inferred for QSO–DLA environments but typical of Galactic molecular clouds. These lines, which have not been detected in QSO–DLAs, might have their origin in the vicinity of the GRB (Vreeswijk et al. 2004). For GRB 050730 and GRB 051111, two excitation mechanisms, collisional excitation and indirect UV pumping by the GRB afterglow, were considered by Prochaska, Chen, and Bloom (2006). The radiative pumping by an IR or UV radiation field produced by intense star formation in the GRB environment was also considered by Berger et al. (2005) for GRB 051111. For this burst the detection of strong absorption from Fe\(\text{II}\) and Si\(\text{II}\) fine structure levels led Penprase et al. (2006) to conclude that the local environment of GRB 051111 is warm and dust-depleted. In the case of collisional excitation, they inferred a temperature of about 10\(^3\) K, but Berger et al. (2005) considered that the collisional excitation is unlikely to be the dominant mechanism for this burst. Even if more work is needed to choose between these different scenarios Berger et al. (2005) were convinced that GRB afterglows will allow probing the physical conditions within individual star-forming regions directly and will provide direct information on the conditions of star formation across a wide range of redshifts.

One last point: if the presence of these excited species arises as a result of the intense ionizing radiation of the afterglow, we expect significant temporal variability in the strength of the excited lines. The time-dependence of the absorption features due to the gradual ionization of the surrounding medium by the afterglow radiation has been analyzed by Perna and Loeb (1998). A tentative detection of a temporal change (two epochs ~16 h apart) in the fine structure line of Fe\(\text{II}\) in the spectrum of GRB 020813 has been reported by Dessauges-Zavadsky et al. (2006) with a significant decline by a factor of 5 in the equivalent width of the Fe\(\text{II}\) transition at \(\lambda = 2396\) Å. Further observations are needed to clarify the question of fine structure transitions and to identify the mechanisms for their excitation, which may reveal the physical conditions of the gas where they occur. This short analysis of the possibilities offered by the fine and broad spectroscopy of early GRB afterglows demonstrates the
richness of this field for a better understanding of star formation in the early Universe. Let us now jump to another topic, the study of GRB host galaxies.

### 9.3 HOST GALAXIES

Another way to investigate the GRB environment and the properties of the host galaxies is the study of their emission lines which can be observed when the afterglow has faded away, several days after the burst. This is much easier with GRBs, because in QSO–DLAs the host galaxy is always illuminated by the background quasar. We can summarize the properties of GRB host galaxies since a large sample has been collected with multi-wavelength photometric observations or with emission line spectroscopy.

On average GRB host galaxies have a small size and a low luminosity. They are generally blue with strong emission lines and with active star formation (Sokolov et al. 2001, Le Floc’h et al. 2003, Christensen, Hjorth, & Gorosabel 2004, Gorosabel et al. 2005, Fynbo et al. 2005, Sollerman et al. 2005, Stanek et al. 2006). GRBs appear to trace the blue light of their host galaxies. They are more concentrated in the very brightest regions of their host galaxies than are core collapse SNe, suggesting that GRBs are formed from the most massive stars. Moreover GRB host galaxies are fainter and smaller than supernova hosts. Fruchter et al. (2006) found striking differences between these two galaxy populations. Of 42 long GRB hosts, 41 appear to be small star-forming galaxies with sizes typically like those of the Magellanic clouds (LMC and SMC, with metallicities 0.3 $Z_\odot$ and 0.2 $Z_\odot$, respectively). The morphology of GRB hosts is therefore consistent with the idea that long GRBs occur at lower metallicities than normal core-collapse supernovae (see also Modjaz et al. 2008). Wolf and Podsiadlowski (2007), using the sample of Fruchter et al. (2006), showed that the median GRB host galaxy is a galaxy with the mass of LMC and half-solar metallicity. Moreover, a large majority of GRBs are associated with irregular galaxies. This is another difference with respect to core-collapse SNe, which are equally distributed between spiral and irregular galaxies (Fruchter et al. 2006). These authors concluded that even if core-collapse SNe and long GRBs are formed by the collapse of massive stars, the distribution of these objects within their hosts and the nature of these hosts are substantially different (but this is not the conclusion of Kelly, Kirshner, & Phare (2007) who found that SN Ic occur in environments similar to that of GRB hosts). Fruchter et al. (2006) suggested that GRBs could be associated with the most massive stars and restricted to galaxies of limited chemical evolution (see also Modjaz et al. (2008) for similar conclusions): broad-lined SN Ic without GRB tend to inhabit luminous and more metal-rich galaxies than GRB–SNe, consistent with the hypothesis that low metal abundance is the cause of very massive stars becoming SN–GRBs (see also Stanek et al. 2006). Fruchter et al. (2006) reported the observations of seven GRB hosts with measurements or limits on their metallicity which are all less than one-third solar. They argued that low metallicity seems to play a decisive role in the GRB phenomenon and they suggested an upper limit to the metallicity of $Z \sim 0.15 Z_\odot$ for cosmological
GRBs. This result comes from a correlation they found between host metallicity and isotropic energy $E_{\text{iso}}$, which they interpreted as a metallicity cut-off around $Z \sim 0.15 Z_\odot$. But this relation was not confirmed by Wolf and Podsiadlowski (2007, see their Figure 5). In the same way Savaglio, Glazebrook, and Le Borgne (2008) did not confirm this correlation (their Figure 17). Stanek et al. (2006) studied local GRB hosts ($z < 0.25$), and found that they are faint and metal-poor compared to the population of local star-forming galaxies. Finally, low metallicity has also been invoked to explain why no GRB (of the long type discussed here) had been found in sub-millimeter bright galaxies (Fruchter et al. 2006). These highly dusty galaxies with substantial metallicity are found at redshift $z = 1–3$, with a median redshift of 2.2, and they seem to be the site of a large fraction of massive star formation in the distant Universe (Chapman et al. 2005).

Savaglio (2006) summarized the metallicities of 11 GRB hosts (her Table 3) and found that for a redshift range between 0.01 and 1 the average value of the metallicity was 0.7 times the solar value $Z_\odot$.

The comparison between metallicities obtained through emission line detection in the host galaxies and those derived in GRB–DLAs (discussed in the previous section) is briefly discussed by Savaglio (2006). This question can be best addressed when metallicity is derived using both the absorption lines in the afterglow and the emission lines of the host, noting that the observations of the host galaxy can be made at any time. In any case the metallicity derived in GRB–DLA involves the material along the line of sight, while the metallicity derived from emission lines is a luminosity-weighted average of the galaxy (at least for hosts situated at $z > 0.2$, which represent 85% of the spectroscopy sample considered by Savaglio, Glazebrook, & Le Borgne (2008)). Nevertheless, if it is assumed that this comparison can be made, the comparison of the average metallicity of 11 GRB hosts at $z \leq 1$ (0.7 $Z_\odot$) with the values measured at high redshift in GRB–DLAs (Figure 9.8) suggests that the metallicity has a general decreasing trend beyond $z = 1$.

Finally, Savaglio (2006) compared GRB–DLA observations with the predictions of an empirical model that gives the metallicity of galaxies for various stellar masses as a function of the Hubble time (Savaglio et al. 2005). GRB host galaxies are found to have low stellar masses, in the range $10^{8.6–9.8} M_\odot$, slightly smaller than the LMC. Figure 9.10 shows the stellar mass histogram of 32 GRB hosts, compared with the same histogram obtained for 201 normal galaxies from the Gemini Deep Deep Survey in the redshift range $0.4 < z < 2$ (Abraham et al. 2004b). Of the GRB hosts 79% have masses below $10^{10} M_\odot$. The median mass is $10^{9.5} M_\odot$, comparable to the mass of the stellar content of the Large Magellanic Cloud. Therefore, at high redshift GRBs seem to be tracing lower-mass galaxies than traditional high-$z$ surveys. A similar result was obtained by Chary, Becklin, and Atmus (2002) for seven GRB hosts.

The optical spectrum of GRB 050904 at $z = 6.295$ indicated that at this high redshift the circum-burst medium is already enriched in metals and that the mass metallicity ($M–Z$) relation obtained for 53 000 local ($z \leq 0.1$) star-forming galaxies (Tremonti et al. 2004) holds up to $z > 6$ (Berger et al. 2007). This relation has been extended to redshifts $0.4 < z < 1$ by Savaglio et al. (2005), and to $z = 2$ by Erb et al.
Figure 9.10. Estimate of the total stellar mass content of 32 GRB host galaxies, based on SED fitting to the multi-band optical/NIR photometry. A comparison is given with 201 normal galaxies between $z = 0.4$ and 2. Of the GRB host sample 79% have stellar masses below $10^{10} M_\odot$ (from Savaglio, Glazebrook, & Le Borgne 2006).

Figure 9.11. Mass–metallicity relation for seven GRBs (filled circles) and normal star-forming galaxies for redshifts between 0.4 and 1. Even though the sample is too limited, GRBs behave like the other galaxies in the same redshift range. The outlier at log $M = 8.8$ is the host galaxy of GRB 991208 at $z = 0.7$ (Castro-Tirado et al. 2001). The straight line is the bisector fit for the sample of galaxies (from Savaglio, Glazebrook, & Le Borgne 2006).

As an example, Figure 9.11 gives the mass–metallicity relation observed in normal star-forming galaxies with $0.4 < z < 1$. Even though the sample is limited, the seven GRB hosts behave like the other galaxies in the same redshift range. Savaglio, Glazebrook, and Le Borgne (2006) concluded that there is no evidence that GRB hosts represent a population different from typical galaxies found in the young Universe. The fact that GRBs are more likely to be found in faint galaxies is expected since these galaxies are the most common ones in the Universe at low redshift, as well as at high redshift. The same effect exists for galaxies hosting type Ia SNe. This is not, however, the conclusion of Langer and Norman (2006) or of Stanek et al. (2006). These authors argued that GRB hosts are more likely to be metal-poor, low-mass galaxies, because massive metal-poor stars (the GRB progenitors) are rare in massive galaxies.
To estimate the star formation rate (SFR) of 19 GRB host galaxies, Savaglio, Glazebrook, and Le Borgne (2006) used the O\textsc{ii} emission, which is the most common line measured in GRB hosts. They found a median dust-corrected SFR of 12 $M_\odot$ yr$^{-1}$ (assuming a Milky Way extinction curve with $A_V = 1$). The range spanned by the 19 hosts is between 1 and 100 $M_\odot$ yr$^{-1}$. Given the generally low stellar masses of these GRB hosts, they concluded that a large fraction of them are undergoing a burst of star formation. In a study of 32 GRBs with measured redshifts, Savaglio (2006) reported that at least 56% of the hosts are undergoing a burst of star formation.

This short analysis of the properties of GRB host galaxies gives the opportunity to mention a morphological study of 42 GRB host galaxies performed by Wainwright, Berger, and Penprase (2007). Among the general properties that are established, they note the high fraction (30%) of merging systems, showing clear signs of interaction, and an additional 30% showing irregular and asymmetric structure. These characteristics seem to be independent of the redshift and of the galaxy luminosity. Hence GRBs might be over-represented in faint merging systems which would be the sites of intense star-formation activity. In conclusion, it is fairly clear that GRB hosts are faint galaxies with low stellar mass. GRBs seem to occur in metal-rich and metal-poor regions, and with only a weak dependence of the metallicity on the redshift. Fryer et al. (2007) presented an interesting discussion of this metallicity issue. They concluded that there is no consistent picture of the metallicities of GRBs and they noted that direct measurements (absorption and emission line spectroscopy) argue for higher metallicities, whereas indirect measurements (interstellar extinction in the host and host morphology) suggest lower metallicities. Today, the question of the metallicity of GRB hosts remains an open issue.

Another important issue is the dust in GRB hosts, whose extinction properties are still poorly known. If dust is very abundant along the GRB line of sight it is not clear why many GRBs show very little reddening. As we have already noted, clumpy dust might be present (Witt & Gordon 2000), but as for the metallicity, low dust contents have also been reported. Finally, with low dust content, low metallicity, and a chemical composition enriched in $\alpha$-elements, the GRB hosts have characteristics similar to those of DLA systems at $z > 2$ (Chen, Prochaska, & Bloom 2006).

These issues have been reviewed by Savaglio, Glazebrook, and Le Borgne (2008) for a large sample of GRB host galaxies (46 objects) in the redshift range 0 < $z$ < 6.3, with a mean redshift $z = 0.96$ (89% of the hosts being at $z < 1.6$). We give here their conclusions concerning the main characteristics of the GRB hosts, as a complement to our previous analyses. The average host stellar mass is $10^{9.3} M_\odot$. The average metallicity for a subsample of 17 hosts is about 0.25 $Z_\odot$. Metallicities derived from emission lines in the host galaxies at $z < 1$ are relatively low, around 0.1 $Z_\odot$, comparable with metallicities obtained from UV absorption lines in the cold medium of GRB hosts at $z > 2$ (GRB–DLA). Combining these two results, it can be said that no significant evolution of metallicity seems to be present in the GRB hosts in the range 0 < $z$ < 6. The dust extinction in the visual band is $A_V = 0.5$. The star-formation rate (SFR) is in the range 0.01–36 $M_\odot$ yr$^{-1}$, with a median specific SFR (SFR per unit stellar mass) of 0.8 Gyr$^{-1}$ (see their Tables 9 and 11 giving a very complete summary and properties of GRB host galaxies). According to their analysis, most
GRBs are associated with the deaths of young massive stars and there is no compelling evidence that GRBs hosts are peculiar galaxies. Of course these conclusions do not apply to the subclass of short hard GRBs, which need more data to understand the nature of their host galaxies.

To summarize we can say that the GRB environment is characterized by large neutral gas column density, often with low metallicity and low dust content. The presence of strong fine structure transitions indicates an extreme environment in terms of gas density and temperature, rarely observed in classical DLAs. But metal enrichment around GRBs and dust content in the star-forming regions where GRBs are born are important issues which still require study to draw robust conclusions. Despite these uncertainties and the divergences we have mentioned, the studies of GRB afterglows and host galaxies provide a new way to probe the chemical state of the young Universe, and they might soon open a new window on the formation of the first generation of stars. These studies require a large sample of high-z GRBs with early afterglow observations in the NIR; they rely on the spectroscopy of bright afterglows undertaken quickly after the burst.

**Cosmic rays, neutrinos, gravitational waves**

Let us now go to the second part of this chapter, devoted to the emission of very high-energy photons and non-electromagnetic signatures which accompany or might accompany GRBs. This part is short for two reasons: it is very speculative with few results on very high-energy photons (greater than tens of gigaelectronvolts) and no detection of non-electromagnetic emissions. The second reason is the existence of complete and well-referenced reviews by Zhang and Mészáros (2004) and by Mészáros (2006). We will mention here what we consider to be the most relevant points, knowing in advance that we will not be exhaustive on these topics.

### 9.4 HIGH-ENERGY PHOTONS

#### 9.4.1 Brief summary of high-energy gamma-ray observations

The detection of photons in the megaelectronvolt range (MeV) has been one characteristic of the GRB prompt emission since the beginning of the long GRB story. Before the launch of the Compton Gamma Ray Observatory (CGRO), the Gamma-Ray Spectrometer on-board the Solar Maximum Mission (SMM, e.g. Matz et al. 1985, Harris & Share 1998), and the French-Soviet Phèbus detectors on the Granat mission (Barat et al. 1992, 2000, Barat & Lestrade 2007) detected GRBs up to 10 MeV. They showed that very often GRB spectra extend to these relatively large photon energies. On the other hand, emissions in the 10 to 100 MeV energy range have only been reported for a limited number of GRBs, mainly with Comptel and Egret on CGRO (Matz et al. 1995, Hanlon et al. 1994, Winkler et al. 1995, Kippen et al. 1998). Comptel observed 29 GRBs in the range 0.75–30 MeV in 4 years. While a
break in the BATSE energy range is often observable around a few hundred kiloelectronvolts, Kippen et al. (1998) showed that the combination of BATSE, Comptel, and Egret data suggests that, beyond this break, GRB spectra have a power-law emission extending to hundreds of megaelectronvolts.

At higher energies, in the gigaelectronvolt range (GeV), very few detections have been reported: Egret (30 MeV–30 GeV) detected seven GRBs with MeV to GeV emissions (Schneid et al. 1992, Kwoke et al. 1993, Hurley et al. 1994, Sommer et al. 1994, Dingus 1995). GRB 940217 is a particularly interesting burst, which presented a delayed gigaelectronvolt emission (lasting for 90 min!) while the low-energy emission observed with Ulysses (25–150 keV) lasted only 180 s. In this burst an 18-GeV photon was observed 4500 s after the low-energy emission had ended (Hurley et al. 1994). Sommer et al. (1994) reported the detection with Egret of 16 γ-ray photons with energies larger than 30 MeV in GRB 940131, two of them having energies of approximately 1 GeV. The high-energy spectrum can be fitted with a single power law between 30 MeV and 1 GeV. Another event, GRB 941017, was also observed with the calorimeter of Egret between 1 and 200 MeV, simultaneously with BATSE. The spectrum presented a high-energy component (Gonzalez et al. 2003; Figure 9.12) which was not the power-law extension of the emission detected below 1 MeV by BATSE’s large-area detectors. As shown in the figure the high-energy component decayed more slowly than the lower-energy component and the shape of the spectrum led the authors to claim that this high-energy component was inconsistent with the synchrotron shock model.

In the teraelectronvolt range (TeV), many ground-based Čerenkov detectors have participated in the search for emission from quiescent sources and GRBs. Milagrito was a wide-field-of-view teraelectronvolt observatory using a water Čerenkov detector of size 35 × 55 × 2 m located at an altitude of 2650 m near Los Alamos. It contained 723 photomultiplier tubes immersed in a large water reservoir (24 × 10⁶ liters, Atkins et al. 1999, McCullough 1999). Milagrito operated in 1997–1998 and was then replaced by Milagro which has a better sensitivity at low energies (between 0.1 and 10 TeV). Fifty-four GRBs were detected by BATSE during Milagrito activity, but only one, GRB 970417A, was associated with an excess of counts in Milagrito (Atkins et al. 2000). The excess had a chance probability of 2.8 × 10⁻⁵ of being a background fluctuation, but this probability reduces to 1.5 × 10⁻³ if one takes into account the fact that 54 GRBs were searched (Atkins et al. 2005, Abdo et al. 2007). If this association is real, the fluence above a few hundred gigaelectronvolts would be at least an order of magnitude larger than the sub-megaelectronvolt fluence. Due to the opacity of the IGM at such high energies, the source had to be a nearby object (z < 0.3). This detection was not confirmed by the GRAND muon detector array of the University of Notre Dame (USA), which detects muons in air showers, and can detect gamma-rays (0.01–1 TeV) through the secondary muons produced in the gamma-ray-hadronic cascade in the atmosphere with pions decaying to muons. But GRAND may have detected sub-teraelectronvolt emission from GRB 971110 (Poirier et al. 2003), among the eight GRB candidates which were observed. The muon event, coincident in time and direction with the BATSE detection, has, however, a low significance (2.7σ). An extrapolation of the BATSE γ-ray spectrum would
Figure 9.12. Evolution of the spectrum of GRB 941017 as a function of time. Five consecutive time intervals have been selected (Gonzalez et al. 2003). The measurements of the LAD detectors of BATSE and the TASC of Egret are indicated with crosses and filled circles respectively. Solid curves show model fits to the data. The two spectral components (dashed lines) and their time evolution are described at low energies by the Band function and at higher energies by a power law (from Gonzalez et al. 2003).
lead to a very low muon signal, compared to the excess of $466 \pm 171$ muons reported. Between October 1995 and March 1999, the data of the Tibet Air Shower Array were compared with BATSE observations of 69 GRBs, in a search for coincident multi-teraelectronvolt signals (Amenomori et al. 2001). No significant counterpart was detected.

In summary, we have very few results on the very high-energy (TeV) emission from GRBs. Nevertheless, Fragile et al. (2004) attempted to interpret the possible detections, not claimed as definitive, in the context of three models which, as we will see later, have been proposed to explain the production of teraelectronvolt $\gamma$-rays: inverse Compton scattering (ICS) of ambient photons by relativistic electrons in the burst environment, proton synchrotron emission (Vietri 1997, Totani 2000), and inelastic scattering of relativistic protons off ambient photons, creating high-energy neutral pions $\pi^0$ decaying into $\gamma$-rays (Waxman 1995, Waxman & Bahcall 1997). For these authors ICS and proton synchrotron appear to be the most efficient processes to produce teraelectronvolt $\gamma$-rays.

9.4.2 High-energy gamma-ray observatories

At these very high energies, the development of many ground-based air shower and water Čerenkov telescopes (planned or already working) is very promising and will certainly put interesting constraints on GRB models in the coming years. Let us say a few words about these new-generation telescopes with the following non-exhaustive list.

(a) Cangaroo III is an array of four 10-m atmospheric Čerenkov telescopes which has been operational since 2004. It explores the $\gamma$-ray sky above 0.1 TeV with an angular resolution better than 0.1° (Kubo et al. 2004).

(b) The Magic telescope (Major Atmospheric Gamma-ray Imaging Čerenkov Telescope, 17 m in diameter) is installed at the former Hegra site on the Canary Island of La Palma. The parabolic tessellated mirror has an area of $234 \text{ m}^2$ with 943 $50 \times 50 \text{ cm}^2$ individual mirror elements (Lorenz 2004). The telescope was completed in 2004.

(c) Veritas (Very Energetic Radiation Imaging Telescope Array System) is a new-generation observatory extending the Whipple 10-m Telescope (Weekes et al. 2002, Horan et al. 2005). Veritas is an array of four large imaging telescopes operating in the range 0.2–1 TeV, with a field of view of 3.5°. Weekes et al. (2002) compared the expected sensitivity of Veritas with the sensitivities of other telescopes operating in the same energy range (Whipple, Magic, Egret, Milagro and with the past and future space telescopes: Egret and GLAST; Figure 9.13). The sensitivity of Veritas is $\sim0.02 \text{ Crab}$ above 300 GeV for a 50-h exposure. The full array with four telescopes was completed during the summer of 2007.

(d) HESS (High Energy Stereoscopic System) is a system of four 107-m² mirror area imaging Čerenkov telescopes (Hinton 2004) which is now completed and operates in Namibia. The threshold is low, 100 GeV, and its sensitivity is
0.01 Crab for a 50-h exposure (comparable with Veritas). Given the successful operation of HESS (which was awarded the Descartes prize of the European Community in 2007 for its scientific achievements), it was decided to extend the HESS array with a fifth telescope with a larger area of 600 m² and a powerful camera made up of 2048 photomultiplier tubes, which will lower the energy threshold down to 20 GeV. This project, called HESS-2, will be completed in 2008.

There are also water or ice Čerenkov telescopes:

(e) *Amanda* (Antarctic Muon And Neutrino Detector Array) consisted of 677 photomultiplier tubes housed in optical modules placed beneath the surface of the ice at the South Pole. Its threshold was 50 GeV (Ahrens et al. 2003). Especially designed for high-energy neutrino detection, it could also detect GRBs. Four hundred GRBs occurred during the 7 years of Amanda observations, and simultaneous neutrino excesses have been searched in the data without positive detection (Achterberg et al. 2008).

(f) *IceCube* represents the new generation after Amanda. It consists of 1 km³ of Antarctic ice instrumented with 4800 photomultiplier tubes (Karle et al. 2003, Karle 2006).

(g) *Milagro* is a water Čerenkov detector sensitive to γ-rays above ~100 GeV with a detector consisting of a 6-million-gallon artificial pond with 723 photomultiplier

These ground-based telescopes are completed by two space missions which will explore the sky in the energy range 30 MeV–300 GeV. The smaller of the two, AGILE (Tavani et al. 2006), a mission led by the Italian Space Agency, was launched on April 23, 2007. GLAST (Ritz, Michelson, & Meegan 2007, and references therein) is a more ambitious mission led by NASA, which was successfully launched on June 11, 2008 and renamed Fermi. These missions, as we will see in the following paragraphs, will undoubtedly clarify many issues which remain open in spite of the enormous progress already made by BeppoSAX, HETE-2, and Swift, in conjunction with multi-wavelength and quasi-simultaneous observations by the largest telescopes on the ground and in space.

9.4.3 High-energy gamma-ray emission from internal and external shocks

We discuss now the processes and sites which can produce very high-energy photons. Let us concentrate first on VHE photons with giga- to teraelectronvolt energies, which may be produced by various emission processes operating in several emission sites. These processes can involve the leptonic component (electrons) of the fireball, for instance inverse Compton scattering in external or internal shocks; they can also involve the hadronic component (protons and neutrons), for instance p–γ interactions with pion production or proton synchrotron radiation. In baryonic outflows, neutrons are likely to be present; when the neutron component decouples from the proton component, p–n inelastic collisions lead to pions, among them π°, which produce UHE (ultra-high-energy) photons, which cascade down to the gigaelectronvolt range (Derishev, Kocharovsky, & Kocharovsky 1999, Bahcall & Mészáros 2000, Rossi, Beloborodov, & Rees 2006). A full discussion of the different processes which can produce these VHE and UHE photons is given by Zhang and Mészáros (2004). They reviewed several mechanisms operating at various possible sites capable of producing high-energy photons from a relativistic fireball:

- Electron self-IC component from the external reverse shock, or cross-IC components between the reverse and the forward shocks (Mészáros & Rees 1994, Wang, Dai, & Lu 2001a,b).
- Proton synchrotron emission and photo-meson cascade emission from the internal shocks (Rachen & Mészáros 1998).
• Cascade emission resulting from $p$–$n$ inelastic collisions during the early phase of fireball evolution after the decoupling phase (Derishev, Kocharovsky, & Kocharovsky 1999, Bahcall & Mészáros 2000).

We will consider below a few examples in greater detail, illustrating the various possibilities for producing a high energy $\gamma$-ray component. New high-energy data on GRBs might help to constrain some of the fireball shock parameters, particularly $\varepsilon_e$ and $\varepsilon_B$ (the energy density carried respectively by the electrons and the magnetic field), whose values are not well known.

9.4.4 High-energy gamma-ray emission involving external shocks

**Electron self-IC component from the external forward shock**

Our first example comes from Zhang and Mészáros (2001). These authors investigate the proton synchrotron and the electron Inverse Compton (IC) emissions and their possible signatures in the broadband spectra and in the high-energy light-curves of GRB afterglows (keV–GeV). In the $\varepsilon_e$–$\varepsilon_B$ parameter space they identify the regions where the various radiation mechanisms they consider are dominating (Figure 9.14). Depending on the frequency range (below or above the electron–synchrotron cutoff), the competition between the various components is analyzed: electron synchrotron, proton synchrotron, electron IC, and hadron-related photo-meson decay components (see the complete discussion given by Zhang & Mészáros, 2001). Their conclusion is that the most likely origin for an extended high-energy afterglow component at giga-electronvolt energies (similar to the one found in GRB 940217) is from the electron IC component. The detectability of this IC component is favored by a high-density external medium. The electron IC component will peak at giga-electronvolt energies and should be detected by GLAST. They gave the giga-electronvolt fluence versus time compared with the Egret and GLAST sensitivities for three types of bursts (Figure 9.15). For the IC-dominated regime, an extended (1 h long) giga-electronvolt afterglow should be detectable by GLAST for typical bursts at $z \sim 1$ (but see Zhang (2007a) for a less optimistic prediction).

**Proton synchrotron emission and photo-meson cascade emission from the external shock**

Another mechanism for the production of high-energy photons was originally proposed by Vietri (1997): proton synchrotron emission. Vietri considered the possible production of $10^{20}$ eV cosmic rays and showed that these particles might be the source of photons up to 300 GeV via synchrotron emission as they pass through the acceleration region. These high-energy photons carry away a small fraction ($\sim$0.01) of the total burst energy. In their model the interactions are in the afterglow, the
Figure 9.14. Regions in the $\varepsilon_e, \varepsilon_B$ parameter space in which the various radiation mechanisms dominate at selected frequencies. For photon energies $\nu < \nu_{\text{IC}}$ (the synchrotron frequency for electrons at the upper end of their energy distribution) lines 1 and 2 (solid lines) divide the space into three regimes. In regime I the proton synchrotron component overcomes the electron synchrotron component. In regime II the electron IC component overcomes the electron synchrotron component. In regime III the electron synchrotron component dominates the other two. For a higher energy band, with $\nu > \nu_{\text{IC}}$, the space is divided into regimes I and I' by line 3, or line 4 or 4' (dashed lines), depending on the subcase. In regime I' the proton-related components overcome the electron IC component, while in regime II' the electron IC dominates over the proton components. For $\nu < \nu_{\text{IC}}$ (the IC-boosted frequency of synchrotron photons radiated by electrons at the cooling break energy), the separation is given by line 3, which does not depend on the frequency. For $\nu > \nu_{\text{IC}}$ the separation line is frequency-dependent, and given by line 4 or 4', which is drawn for $\nu \sim 10^{26}$ Hz. Line 4 assumes that the reduction factor is $k = 0.1$, while line 4' assumes $k = 1$ (see Zhang & Mészáros 2001). The dependences of the separation lines on $t$, $n$, and $E$ are indicated on the plot, which cause the different regimes to enlarge or shrink with these parameters (from Zhang & Mészáros 2001).

Photon energy density is lower (compared with the internal shock case proposed, for instance, by Waxman & Bahcall (1997)) and there is no energy degradation in situ. Nevertheless for the 300-GeV photons, $\gamma\gamma$ interactions in the IGM will degrade the energy of the photons, reducing the observability distance to $\sim 300$ Mpc ($z \leq 0.1$). A difficulty with this model was pointed out by Gallant and Achterberg (1999): the production of UHECRs by the blastwave expanding into the ISM would be impossible. They showed that, with Fermi-type shock acceleration, initially isotropic particles in the upstream medium can gain a factor of $\sim \Gamma^2$ in energy in the first shock
crossing cycle but only a factor of 2 in subsequent shocks, while Vietri (1997) assumed that the energy of the particle (proton) is multiplied by $\Gamma^2$ in each shock crossing. Under these conditions the maximum energy attainable is well below $10^{20}$ eV. To avoid this difficulty, Gallant and Achterberg (1999) proposed that the blastwave expands into a pulsar wind bubble (Königl et al. 2002). Pulsars emit relativistic winds which must also contain ions. The relativistic flow of the pulsar wind is thermalized in a termination shock and this shocked material forms a relativistic plasma bubble in the interstellar medium. The ions approximately conserve their post-shock energy throughout the bubble. Gallant and Achterberg (1999) determined the typical energy of the thermalized post-shock ions. This energy may be high: $m_i\gamma_w c^2$, with $\gamma_w$ the bulk Lorentz factor of the pulsar wind. The typical energy of these relativistic ions is boosted by the blastwave when it is in the free expansion phase: $E_{\text{max}} \sim 2\eta^2 m_i\gamma_w c^2$, with $\Gamma \sim \eta \sqrt{2}$, $\eta = E_{\text{GRB}}/Mc^2$, and $M$ is the baryonic mass in which the fireball energy $E_{\text{GRB}}$ is initially deposited (see Chapter 5). Under these conditions the ions can

Figure 9.15. Gigaelectronvolt fluence versus time for three types of bursts starting from $t_{\text{dec}}$ and ending at the time when the bulk Lorentz factor $\Gamma = 2$. The energy flux has been integrated within the range 400 MeV to 200 GeV in order to compare with observations. The solid curves I, II, and III indicate bursts in the three regimes considered in Zhang and Mészáros (2001; see Figure 9.14), where the emission is respectively dominated by proton synchrotron emission, electron IC emission or electron synchrotron emission. GRBs are assumed to be at redshift $z = 1$. The case of GRBs at a redshift $z = 0.1$ is illustrated by the three dotted and unmarked curves. The sensitivities of GLAST and Egret are also reported (from Zhang & Mészáros 2001).
reach ultra-high energies, but this assumes the presence of pre-existing relativistic particles of sufficient energy to be boosted in the first shock-crossing cycle (Gallant & Achterberg 1999). This is not at all the usual condition of fireballs expanding into a typical interstellar medium. In this last case, the highest energies which can be produced by the blastwaves seem to fall well below the domain of the UHECRs. In conclusion, the question of UHECR production in GRBs is really not trivial (see Section 9.6).

Another calculation of the production of high-energy $\gamma$-rays from UHECR protons in the blastwave associated with a GRB was proposed by Böttcher and Dermer (1998). The conversion of kinetic energy into radiation occurs when the blastwave decelerates as it sweeps up matter from the external medium. Protons and electrons are accelerated, since the blastwave can accelerate protons to $\sim 10^{20}$ eV (see, however, the previous remarks from Gallant & Achterberg (1999)). Böttcher and Dermer (1998) computed the photon and positron spectra resulting from decaying pions produced by UHECR interacting via photo-meson interactions with photons. They calculated the emergent synchrotron radiation from protons, from positrons (coming from $\pi^+$) and from pairs produced by $\gamma-\gamma$ interactions. The high-energy $\gamma$-ray radiation from proton synchrotron and from the photo-pion induced pair cascades decays more slowly than the low-energy afterglow and the Synchrotron Self Compton (SSC) radiation because the cooling of the electrons is more efficient than the cooling of protons. According to Böttcher and Dermer (1998), an increasing ratio of the high-energy $\gamma$-ray flux to the optical or soft X-ray flux in the afterglow would be evidence for proton synchrotron radiation and for UHECR acceleration in GRBs. The predicted flux between 10 MeV and 100 GeV could be detected by GLAST and by ground-based air Čerenkov telescopes with low energy thresholds.

However, according to Zhang and Mészáros (2001), the proton synchrotron emission and the hadron related photo-meson electromagnetic components are likely to be less important in the afterglow than previously estimated (Vietri 1997, Böttcher & Dermer 1998, Totani 2000). In the $\varepsilon_e - \varepsilon_B$ phase space, the region in which this proton synchrotron component overcomes the electron synchrotron and IC components is quite small but non-negligible (region I in Figure 9.14). It is characterized by $\varepsilon_e \ll \varepsilon_B$ (unless the medium density is high). Zhang and Mészáros (2001) analyzed the detectability of the proton synchrotron component, even though it is less probable, since the detection of this component would be very important for constraining the fireball and shock parameters. As far as $\gamma\gamma$ absorption is concerned, Zhang and Mészáros (2001) found that it is not important for the external shock below the teraelectronvolt range.

Pe’er and Waxman (2005) have also considered high-energy photon emission in GRB early afterglows. They took into account electron synchrotron emission, inverse Compton scattering, $e^+/e^-$ pair production and the photo-production of pions.$^3$

$^3$ High-energy baryons produce energetic pions (via photo-meson interactions with the low-energy photons) which are sources of high-energy photons and neutrinos (see Section 9.6, Böttcher & Dermer 1998, Waxman & Bahcall 1997, 2000).
with the resulting electromagnetic cascades. They found a dependence of the spectrum on the ambient density and on the magnetic field energy fraction which is pronounced at high photon energies. As an example they obtained a comparable flux at 1 GeV and 1 TeV if the density is low and $\varepsilon_B \sim 10^{-4}$, while a large flux ratio between these two energies is obtained for a dense medium and $\varepsilon_B$ near equipartition. They also discussed the respective roles of synchrotron emission from shock-accelerated protons and the photo-production of pions to explain teraelectronvolt emission in GRBs (see the possible detection by Milagrito of teraelectronvolt photons from GRB 970417A discussed above; Atkins et al. 2000). To explain this event Totani (1998a,b) had proposed proton synchrotron emission. An alternate explanation was preferred by Pe’er and Waxman (2005) because the explanation of Totani required that electrons carry a very low fraction of the energy ($\varepsilon_e \sim 10^{-3}$) whereas $\varepsilon_e$ seems to be near equipartition in the afterglow. Pe’er and Waxman (2005) preferred the photo-production of pions (and pion decay) which may become comparable to that of the IC emission of shock-accelerated electrons when the fireball expansion occurs into a high-density wind. With their model Pe’er and Waxman (2005) predicted the emission of high-energy photons, between $\sim 1$ GeV and a few teraelectronvolts. The 1–10 GeV predicted flux is within the detection capabilities of GLAST and the sub-teraelectronvolt emission is detectable by the new generation of ground Čerenkov telescopes. At 1 TeV they expect the detection of $\sim 1$ burst per year from low-redshift ($z < 0.3$) GRBs.

Until now we have considered studies implying the afterglows, but to illustrate the large diversity of possible processes and sites we report below some contributions which concern the production of high-energy $\gamma$-rays during the prompt emission, associated with the reverse shock or in connection with X-ray flares.

### 9.4.5 High-energy gamma-ray emission involving internal shocks, reverse shocks or X-ray flares

Pe’er and Waxman (2004) gave detailed calculations of the prompt spectrum of GRB emission due to internal shocks. They found that for small comoving compactness ($l' \leq 3$) the emission extends to energies higher than 10 GeV, with a flux detectable by GLAST. The emission spectrum from internal shocks has also been calculated by Pilia and Loeb (1998); it extends from the optical regime up to very high-energy $\gamma$-rays (10 GeV–1 TeV). Wang, Dai, and Lu (2001a,b) have considered the prompt high-energy $\gamma$-ray emission from Synchrotron Self-Compton (SSC) interactions in the reverse shock and more generally using four inverse Compton processes including SSC in the GRB forward and reverse shocks at very early phases. This comparison led to the conclusion that the SSC emission from the reverse shock dominates over other emission processes (in particular the components of the synchrotron and IC emissions from forward shocks) from tens of mega-electronvolts to tens of giga-electronvolts and that it works for a wide range of shock parameters. This mechanism can explain the prompt high-energy $\gamma$-ray emission from some bright bursts seen by Egret. At teraelectronvolt energies, the combined IC and SSC emissions from the forward shock would dominate over the SSC from the reverse shock, but this depends on the index of the electron power-law spectrum.
Another example is provided by the calculations of Derishev, Kocharovsky, and Kocharovsky (2001). They showed that the prompt emission at sub-megaelectronvolt energies must be generated by the synchrotron mechanism and that the IC scattering process plays a significant role since about 10% of the total GRB energy should be converted into high-energy γ-rays (>100 GeV). In the internal shock model the maximum of the spectrum of the comptonized radiation is model-dependent: it can be as low as 10–100 GeV or as high as 10^3 TeV. In the external shock model it would be 1–100 TeV. The prevalence of synchrotron emission imposes strict limits on the burst parameters, called the SSC constraint by the authors. These constraints come from the fact that radiatively efficient bursts cannot have arbitrary small duration or arbitrary fast variability. The authors also estimate an upper limit of 10^3–10^4 on the Lorentz factor in order to have a high radiative efficiency.

A last example involves IC emission which can be associated with the X-ray flares recently discovered by Swift. As the delayed X-ray flares (up to 10^3–10^4 s after the burst) can overlap the afterglow, IC emission can arise due to the photons of the flare interacting with the electrons in the forward shock. Wáng, Li, and Mészáros (2006) found that this IC emission would produce GeV–TeV flares which might be detected by GLAST and by ground-based teraelectronvolt telescopes. These IC interactions would suppress the synchrotron emission of the afterglow during the flare period, providing an effect which might be detectable. The magnetic field in the afterglow shock might be constrained by the detection of GeV–TeV flares combined with low-energy observations. The authors propose to explain with this process the delayed gigaelectronvolt emission seen by Egret in GRB 940217 (Hurley et al. 1994). Moreover, they argue that, in an external shock scenario for the X-flares, GeV–TeV flares could also be produced by SSC emission of X-ray flare photons. These two scenarios lead to different temporal behaviors of the GeV–TeV flares, which could be used to determine whether the X-ray flares originate from the late activity of the central engine or from an external shock (Wang, Li, & Mészáros 2006).

To close this short list of the various sites and processes which might produce very high-energy photons (beyond tens of gigaelectronvolts) we will discuss now the role of neutrons in relativistic outflows and in neutron-rich outflows.

9.4.6 The role of neutrons

Derishev et al. (1999) have studied the role of neutrons in a relativistic wind composed of neutrons and protons (with typical neutron to proton number ratio ~1). In such winds, the velocities of the neutron and proton flows decouple because the neutrons are not sensitive to electromagnetic forces. This may give rise to an electromagnetic cascade induced by pion production in inelastic collisions of nucleons. This leads to a burst of high-energy photons (coming from π^0 decays) and neutrinos (coming from charged pions π^+ and π^-); a huge number of e^+/e^- pairs is also produced. A fraction of the photons coming from π^0 decays, with energy of 70 MeV, is unprocessed and they appear in the observer’s frame at energies around 100 GeV (due to the high Lorentz factor of the ejecta). They constitute a unique signature of the neutron flow, decoupled from the proton flow, and they give direct information about the Lorentz factor of the fireball. Thus the detection of a burst of 100-GeV
photons by ground-based telescopes promises a better understanding of the physical conditions in the GRB central engine. Bahcall and Mészáros (2000) have also considered this process for the production of high-energy neutrinos and photons (tens of gigaelectronvolts).

Neutron-rich outflows have been studied by Rossi, Beloborodov, and Rees (2006), and Razzaque and Mészáros (2006). The latter authors consider the case of neutron-rich outflows associated with the jets of short GRBs. The jet is initially loaded with neutrons and few protons are expected from neutron star material. Short GRBs are assumed to be produced by the coalescence of two compact objects (BH + NS or NS + NS). They studied the MeV–GeV emission from the neutron-rich relativistic jet. Above the pion production threshold (~140 MeV), the n–p decoupling is followed by inelastic n–p scattering leading to charged pions and \( \pi^0 \). The observed energy of the decay products is ~60 GeV. Among the different components, we only consider here the \( \pi^0 \) decay component. They concluded that for short GRBs at \( z \sim 0.1 \), the photons from the pion decay, at ~60 GeV, should be detectable by large-area Čerenkov detectors such as Milagro. They estimated a rate of about five bursts per year. In addition, these photons, created below the photosphere, would give a gigaelectronvolt component detectable by GLAST, contemporaneous with the usual megaelectronvolt emission of short GRBs. The identification of these components would be quite useful for settling the nature of the progenitors of short GRBs.

### 9.4.7 The interaction of high-energy gamma-rays in the source

As ultra-high-energy photons extend beyond the gigaelectronvolt range, \( \gamma-\gamma \) absorption will introduce a high-energy photon cut-off in GRB spectra. In the region of production of the prompt GRB, the compactness parameter (the optical depth to \( \gamma-\gamma \) absorption) has a decisive role. This optical depth is a strong function of the comoving photon density and of the energy spectrum of the photons in the emitting region. Lithwick and Sari (2001) have used these properties to estimate the lower limits on \( \Gamma \) using a number of bursts observed by EGRET. They considered two main limits based on the requirement that GRBs are optically thin: a limit due to the annihilation of photon pairs and another due to the scattering of photons by pair-created electrons and positrons (the effect of the Compton scattering of photons by the electrons that accompany baryons is not significant). The first of the two limits provides constraints on \( \Gamma \), which are in general between 100 and 400, with very few exceptions. In the near future GLAST observations might lead to the discovery of spectral cut-offs in high-energy spectra which would be very useful to improve these lower limits (Baring 2006). This would complete the values derived from the early optical/NIR afterglows of two GRBs (Molinari et al. 2007; see Section 9.8.4) and various reports of upper limits.

In internal shocks, the simultaneous presence of high-energy photons and low-energy photons has to produce pairs whose secondary emission also contributes to the observed spectrum (Zhang 2007a, and references therein). For instance, Baring
(2006) focused on how attenuation of photons by pair creation internal to the source generates distinctive spectral signatures. Such signatures were not evident in Egret data. They are awaited from GLAST, which is about 10 times more sensitive than Egret, unless the bulk Lorentz factor of the burst emission region is higher than currently argued (e.g. $\Gamma > 10^3$). Razzaque, Mészáros, and Zhang (2004) have also pointed out that, above hundreds of gigaelectronvolts, photons arising in GRB internal shocks are absorbed in the source due to interactions with fireball photons leading to electron–positron pairs. These UHE photons would be absent from the GRB spectrum, typically above 10–100 GeV (see also Lithwick & Sari 2001, Dai & Lu 2002). Thus it is quite reasonable to think that gigaelectronvolt observations will provide interesting constraints on the bulk Lorentz factor (constant or time-dependent) of the GRB outflow through the determinations of the energy of spectral breaks and of the measurements of spectral indices. Razzaque, Mészáros, and Zhang (2004) noted that Baring (2006) did not consider the effects of synchrotron self-absorption, as they did. This inclusion has consequences on the spectrum that emerges from the GRB internal shock region. Besides the expected high-energy cut-off due to pair production, it appears a re-emergence of the spectrum at higher energies (petaelectronvolt energies, PeV), where the shock becomes optically thin to $\gamma-\gamma$ interactions. This is due to the reduction of the spectral density of source target photons. Another point outlined by these authors is the possibility for the fireball to be optically thin to internal $\gamma-\gamma$ interactions at all energies if $\Gamma$ is high enough (typically $\Gamma > 800$).

Let us now consider the absorption of high-energy gamma-rays in the intergalactic medium.

### 9.4.8 The interaction of high-energy gamma-rays with the intergalactic medium

Leaving $\gamma-\gamma$ interactions which take place in the GRB source, let us go to another type of interaction that we have previously mentioned: the interaction of high-energy photons during their propagation between the GRB source and the observer. The attenuation or absorption of high-energy $\gamma$-rays in the intergalactic medium (IGM) has been considered by many groups: Cheng and Cheng (1996), Dai and Lu (2002), and Razzaque, Mészáros, and Zhang (2004). At photon energies ranging from 100 GeV to 100 TeV, pair production with the infrared background radiation dominates. At 100 GeV the mean free path is comparable to the distance of the farthest objects in the Universe, whereas it falls to 10–300 Mpc at 100 TeV, depending on the intensity of the IR background, which is not well known (Plaga 1995). At these energies, even the local Universe is highly opaque (Stecker 1969). Figure 9.16 shows the optical depth of the Universe as a function of redshift (from $z = 0.03$ to 0.3) for high-energy $\gamma$-rays (50 GeV to 100 TeV), as calculated by de Jager and Stecker (2002). This shows that teraelectronvolt sources are strongly absorbed beyond $z \sim 0.1$ ($\sim 300$ Mpc). Pe’er and Waxman (2005) predicted that only about one burst per year can be detected at teraelectronvolt energies (from $z < 0.3$). At 100 GeV, intergalactic absorption is much lower, and 10 bursts per year are expected.
Another interesting effect is the interaction of the very high-energy photons with the cosmic microwave background (CMB). The resulting pairs will Compton scatter on the CMB photons and the microwave radiation will be boosted to MeV–GeV energies. The result of these interactions in the IGM is the appearance of delayed MeV–GeV emission. The delay for the scattered photons to reach the observer would be $10^3$ s or more. For GRB photons of energy greater than 300 GeV, the typical energy of the upscattered CMB photons varies between 50 MeV and 0.8 GeV (Dai & Lu 2002). In their Figure 3 Cheng and Cheng (1996) gave the maximum delay of photons with various energies, assuming a distance of 120 Mpc. This delay is typically $\sim 10^5$ s at 1 GeV and $\sim 10^4$ s at 10 GeV. These delays are also dependent on the strength of the intergalactic magnetic field. Plaga (1995) has considered the role of the magnetic field: he identifies a delayed `after pulse’ of cascade photons arising from IC scattering of electrons and positrons deflected by the IGM magnetic field. These photons are delayed because they have to travel a larger distance to reach the observer than photons coming directly from the source. Plaga (1995) proposed using the high-energy emission from extragalactic sources and GRBs to probe these weak magnetic fields. The delay caused by the scattering of the high-energy photons might be used to measure fields as weak as $10^{-24}$ gauss (Plaga 1995).

Baring (2006) pointed out that internal attenuation can be easily distinguished from extrinsic absorption in the IGM thanks to its distinct temporal behavior. Moreover, the interaction in the IGM would produce attenuation turnovers above 30–50 GeV. The study of these characteristics will be within the possibilities of GLAST and even of ground-based Čerenkov telescopes. Similar studies have been done by Dai and Lu (2002) for teraelectronvolt blazars. Razzaque, Mészáros, and Zhang (2004) computed the secondary spectrum in the 1–100 GeV range, resulting from interactions with the diffuse IR background and with the diffuse microwave
background (CMB). This emission, which comes in addition to the prompt unabsorbed internal shock emission, is delayed by several seconds. Other components can be calculated, for instance the interaction of the secondary γ-rays with the cosmic infrared background, and so on. Finally, these authors mentioned another source of delayed gigaelectronvolt emission, produced in the afterglow: IC upscattering on external shock electrons. This was predicted by Zhang and Mészáros (2001) and would be detectable by GLAST at $z \sim 1$.

To conclude it can be said that high-energy photon emissions would be very useful to have a better description of the IGM while at lower energies (in the energy range of GLAST, 30 MeV–300 GeV), where the attenuation in the IGM is not significant, photon detection will allow putting constraints on the fireball shock parameters and on the emission mechanisms. This short analysis has given the opportunity to illustrate a rich set of processes which can lead to the production of high-energy photons in the internal and external shock regions. The detection of this component (spectra and light-curves) will certainly reduce the large diversity of scenarios which have been proposed to explain the high-energy γ-ray emissions and it may have a major impact on the GRB models. Missions such as AGILE (Pittori, Tavani, & the AGILE Team 2006) and GLAST (Ritz, Michelson, & Meegan 2007, and references therein), launched in 2007 and 2008, will extend the Swift measurements to very high energies and open a new era. Zhang (2007) gave a very interesting list of objectives for GLAST, involving the internal shocks as well as the external forward shocks. Of course these space missions will benefit from the presence of ground-based high-energy photon detectors: water Čerenkov detectors (HAWC, http://hawc.umd.edu/), and atmospheric Čerenkov telescopes (Veritas: Horan et al. 2005, HESS: Hinton 2004, Magic: Lorenz 2004, Cangaroo III: Kubo et al. 2004).

Let us now go to another exciting field: the non-electromagnetic channel, which includes cosmic rays and neutrinos. We will consider first the possibility of producing ultra-high-energy cosmic rays (UHECRs) in GRBs and second the emission of high-energy neutrinos and its role in understanding GRBs.

### 9.5 Ultra-High-Energy Cosmic Rays

In internal–external shock models electrons are accelerated in the shocks, so it is natural to think that baryons are also accelerated by the same shocks. Protons might be accelerated to Lorentz factors up to $<10^{11}$ in the observer frame (Waxman 1995, Vietri 1995) i.e. to energies $E_p < 10^{20}$ eV. This is the domain of UHECRs. The isotropic distribution of UHECRs suggests an extragalactic origin. As they travel through the intergalactic medium they are attenuated by the CMB through photon–meson interactions and scattered by the intergalactic magnetic field. These interactions limit the visibility volume of these CRs to short distances, typically 50–100 Mpc; this is the so-called GZK volume (Greisen 1966, Zatsepin & Kuzmin 1966).

In fact the origin of these UHECRs is still a mystery. Two classes of models have been proposed: in ‘top-down’ scenarios they are the decay products of fossil Grand
Unification defects, whereas in ‘bottom-up’ scenarios they are hadrons accelerated in specific sites such as active galactic nuclei (AGN), and possibly GRBs. Milgrom and Usov (1995) reported the possible association of two of the highest-energy UHECR events with a GRB, and found this coincidence sufficiently remarkable to suggest that the two UHECRs and the GRB could be produced in the same explosion. The best arguments for a potential association of UHECRs with GRBs have been summarized by Waxman (2006b). Waxman showed that, in order to allow proton acceleration up to \( \sim 10^{20} \text{ eV} \), a relativistic wind or fireball must have a Lorentz factor \( \Gamma \) exceeding 100 and a magnetic field energy density \( \varepsilon_B \) close to equipartition. The similarity between these constraints and those imposed on the wind parameters by \( \gamma \)-ray observations (which are based on independent physical considerations), are the basis for the possible association of GRB and UHECR sources. A second argument concerns the energy generation rate of \( \gamma \)-rays by GRBs (\( \sim 10^{44} \text{ erg Mpc}^{-3} \text{ yr}^{-1} \)), which is similar to the energy generation rate of the observed UHECRs (\( 4.5 \pm 1.5 \times 10^{44} \text{ erg Mpc}^{-3} \text{ yr}^{-1} \)), inferred from HiRes Fly’s Eye (Abu-Zayyad et al. 2000) and AGASA (Akeno Giant Air Shower Array; Chiba et al. 1992) measurements.\(^4\) In spite of these arguments, Waxman (2006b) points out that the observed UHECR flux and spectrum can only be accounted for if GRBs inject comparable amounts of energy into the acceleration of high-energy electrons and protons. Unfortunately, the ratio between the energy carried by relativistic electrons and protons in collisionless shocks is not known from first principles. One final comment on these mysterious UHECRs is that their spectrum flattens at \( \sim 10^{19} \text{ eV} \), and there are indications that this spectral change is correlated with a change in composition, from heavy to light nuclei (Bird et al. 1994, Dawson, Meyhandan, & Simpson 1998, Abu-Zayyad et al. 2001). These characteristics would suggest that the CR flux is dominated at energies \( < 10^{19} \text{ eV} \) by a galactic component of heavy nuclei, and at UHE by extragalactic sources of protons. Most of these extragalactic protons of energy \( > 10^{20} \text{ eV} \) would originate from within \( \sim 100 \text{ Mpc} \), and the rate of GRBs out to this distance would be \( \sim 10^{-3} \text{ yr}^{-1} \). The spread in arrival time of protons due to the combined effect of stochastic propagation energy losses and deflection by magnetic fields might be as large as \( 10^7 \text{ yr} \) for a \( 10^{20} \text{ eV} \) proton originating at 90 Mpc. The product of these two quantities leads to the need to have \( \sim 10^4 \) GRBs contributing to the \( > 10^{20} \text{ eV} \) flux (Waxman 2006b).

Let us now address a different question: in the framework of the internal–external shock model, where are the ultra-high-energy protons produced?

Waxman (1995) proposed that type II Fermi acceleration in sub-relativistic internal shocks could account for the acceleration of UHECRs. On the other hand, Vietri (1995, 1998a,b) and Vietri, De Marco, and Guetta (2003) raised critical questions about internal shocks, and proposed that the acceleration of UHECRs is produced at the external shock (see also Böttcher & Dermer 1998). Hence either internal or external shocks, or both, might be involved in the acceleration of UHECRs. However, even though GRBs remain attractive candidates for the production of UHECRs, these scenarios have major caveats as noticed by Zhang and

---

\(^4\) See Nagano and Watson (2000) for a review of the observations and implications of UHECRs. This review includes a discussion of various extensive air shower (EAS) detectors, including AGASA and HiRes.
Mészáros (2004). First, there is no direct proof of the internal shock model, which has been introduced to explain the GRB prompt emission, even if it is the leading scenario. In Poynting-flux-dominated GRB models, for instance, there are no internal shocks and other mechanisms must be invoked to explain the production of UHECRs. The acceleration of protons in the external shock has also been criticized by Stecker (2000) who found that less than 10% of the UHECRs can be accounted for by GRBs, if there is a rough equality between the energy released in the GRB and that released in $10^{20}$ eV cosmic rays. The main reason for this discrepancy is that GRBs are visible out to redshifts larger than 2, while UHECRs are only visible out to $z \sim 0.03$ ($\sim 100$ Mpc), the spatial density of GRBs at low redshift would be too low to produce the observed flux of cosmic rays above $10^{20}$ eV. These arguments were refuted by Vietri, De Marco, and Guetta (2003) who demonstrated that the total energy budget of GRBs is sufficient to account for the energy injection rate of UHECRs as observed by AGASA and HiRes Fly’s Eye. They also argued that the presence of clustered events, doublets, and triplets in the data of AGASA (Takeda et al. 1999, Nagano & Watson 2000) poses no problem for explosive models of UHECR sources, rejecting the objection that impulsive sources like GRBs cannot produce these clustered events. A strong and perhaps more critical objection against an external shock model origin of UHECRs has been given by Gallant and Achterberg (1999) who demonstrated that the acceleration of protons in external shocks is limited to $\sim 10^{16}$ eV for particles which have not been previously boosted by another mechanism (see Section 9.4.3, and Milosavljević & Nakar (2006a), who also do not expect the production of UHECRs in external GRB shocks). Under these conditions the production of UHECRs in GRBs would require a magnetized external medium, as can be found in pulsar wind bubbles (Königl & Granot 2002) or in the supranova scenario (Vietri & Stella 1998).

In view of these questions, only observations can resolve this critical issue of the origin of UHECRs and their possible connection with GRBs. AGNs have long been suspected to be the sources of UHECRs with the acceleration of these cosmic-rays in relativistic jets. The Auger Observatory (Abraham et al. 2004a) might have confirmed this origin. This is the world’s largest cosmic-ray observatory. It includes a 1200-square-mile grid of 1600 large instrumented water tanks which detect particles from air showers. It also includes four sites with a total of 24 telescopes that detect faint fluorescent flashes in the sky, caused by cosmic-ray particles in the air shower. It has been working since 2004. One recent and perhaps major result has been announced in November 2007 (The Pierre Auger Collaboration 2007). Among 27 energetic events detected by Auger above $5.7 \times 10^{19}$ eV, 20 come from the direction of known active galactic nuclei (AGNs). The Pierre Auger scientific team reported that there is less than 1% chance that this correlation between AGNs and UHECRs is random. This result, if it is confirmed with better statistics, will have weighty consequences for the hypothesis that we have discussed: the possible association of UHECRs and GRBs. In any case, there is a possible way to prove that GRBs can produce UHECRs: the search for ultra-high-energy neutrinos associated with GRBs. The emission of high-energy neutrinos from GRBs can be explained only if UHECRs are produced at the GRB site and it would be their signature. So let us now go to the other non-electromagnetic channel: high-energy and ultra-high-energy neutrinos.
9.6 HIGH-ENERGY NEUTRINOS

The presence of accelerated protons at the GRB site must be a reality. Therefore it is reasonable to expect the production of high-energy neutrinos ($\nu$) through various processes $p-\gamma$, $p-p$, or $p-n$ (see for instance Zhang & Mészáros 2004). These authors list the possible production channels of neutrinos in various energy ranges, starting from thermal megelectronvolt neutrinos from the original explosion (Kumar 1999)—which will be extremely difficult to detect from cosmological distances—to petaelectronvolt ($10^{15}$ eV, PeV) and exaelectronvolt ($10^{18}$ eV, EeV) neutrinos. We have seen that GRB fireballs may be neutron-rich (Derishev, Kocbarovsky, & Kochharovsky 1999, Beloborodov 2003). During the fireball acceleration phase, if the initial acceleration is very high, the neutrons decouple from the fireball when the comoving expansion time falls below the nuclear scattering time. The relative velocity between neutrons and protons can be high enough to result in inelastic $p-n$ interactions above the pion production threshold ($\sim 140$ MeV), giving rise to 5–10 GeV neutrinos in the observer’s frame (Mészáros & Rees 2000, Bahcall & Mészáros 2000). These last authors showed that these $\sim 10$ GeV neutrinos could be detectable by km$^3$ detectors if the neutron abundance is comparable to that of protons. These high-energy neutrinos are accompanied by $\sim 10$ GeV photons from $\pi^0$ decay. They might be detectable at low redshift by GLAST. A coincident detection of $\nu$ and $\gamma$-rays would provide constraints on the neutron fraction and therefore on the nature of the GRB progenitor.

In the case of short GRBs, Razzaque and Mészáros (2006) considered a neutron-rich outflow and investigated the emission of high-energy neutrinos ($\sim 50$ GeV) and photons when neutrons and protons decouple. They concluded that the neutrinos expected in their model are unlikely to be detected. Paczynski and Xu (1994) had already suggested that $p-p$ interactions within the internal shocks could also produce $\pi^0$ and give rise to a 30 GeV neutrino burst.

At higher energies, teraelectronvolt neutrinos can be produced within the collapsar scenario. The relativistic jet penetrates the stellar envelope before breaking out and generating the GRB. The internal shocks accelerate protons below the envelope. These protons can interact with thermal photons and protons within the envelope (p–$\gamma$ and p–p interactions), producing pions. A burst of multi-teraelectronvolt neutrinos produced by pion decay could result. For a GRB at $z \sim 1$, it would be easily detectable by a 1 km$^3$ neutrino telescope (Mészáros & Waxman 2001). In a significant fraction of massive stellar collapses, the jet may be unable to punch through the stellar envelope (MacFadyen, Woosley, & Heger 2001), but the teraelectronvolt neutrino signal from such choked jets should be similar to that from ‘successful’ jets, which do break through and lead to observable GRBs. This may provide a means of detecting such choked-off, $\gamma$-dark collapses. Their neutrino emission should be the same as for $\gamma$-ray bright GRBs, but they would be up to 100 times more numerous (Razzaque, Meszáros, & Waxman 2003, Mészáros & Waxman 2001). Another characteristic of these high-energy neutrinos from relativistic buried jets is that their emission starts 10 to 100 s prior to the neutrinos emitted from the GRB fireball in the case of an electromagnetically observed burst. These
neutrinos thus appear as a precursor signal (Mészáros & Waxman 2001). Their flux is shown in Figure 9.17 (Razzaque, Mészáros, & Waxman 2003). The neutrinos from ‘failed GRBs’ could lead to a diffuse background of high-energy neutrinos which could be detectable with IceCube if all core-collapse supernovae produce ‘failed GRBs’ (Razzaque, Mészáros, & Waxman 2003). Detection of this diffuse background of teraelectronvolt neutrinos might be the signature of core-collapse SNe with slower jets, which might contribute to the disruption and ejection of the SN envelope. In this case no γ-ray burst is expected, as the Lorentz factor is a few compared to ~100 for the preceding case of ultra-relativistic jet (Razzaque, Mészáros, & Waxman 2004). Such detections would put interesting constraints on the frequency of γ-ray-
dark’ GRBs from choked jets and hence on the frequency of core collapses of massive stars.

Petaelectronvolt and exaelectronvolt neutrinos can be produced by p–γ interactions within the internal shocks and within the reverse shock and by pion photoproduction respectively (Vietri 1995, 1998b, Waxman & Bahcall 1997, 2000, Guetta, Spada, & Waxman, 2001, Dai Lu 2001, Li, Dai, & Lu 2002, Dermer 2002, Razzaque, Mészáros, & Waxman 2003). Let us study some examples of these very-high-energy neutrino sources. Within the fireball, the internal shocks producing the GRB should also lead to Fermi acceleration of protons to >10^{20} eV (Waxman 1995, Vietri 1995). These accelerated protons might interact with the fireball γ-rays and produce a burst of ~10^{14} eV (100 TeV) neutrinos created by photo-meson production of pions: pγ → Δ^+ → nπ^+ or pπ^0, with π^+ → μ^+ + ν_μ → e^+ + ν_e + ν_μ + anti-ν_μ (Waxman & Bahcall 1997, 2000). This process, which is dominant at the Δ^+ resonance, requires protons with much lower energy than the maximum acceleration energy of 10^{20} eV. Using E_νE_ρ = 0.2Γ^2 GeV^2 to evaluate the energy of protons accelerated at the photo-meson threshold of the Δ^+ resonance, the authors showed that 10^{15} eV protons interacting with ~1 MeV photons (the bulk of the γ-ray energy) may produce a burst of ~10^{14} eV neutrinos (E_ν is the photon energy in the observer’s frame, E_ρ is the energy of the accelerated proton, and Γ is the Lorentz factor of the expanding fireball, typically Γ > 10^2). A km^3 neutrino detector should detect 10 to 100 neutrinos every year, correlated in time and position with GRBs. A few nearby very bright GRBs could even produce a burst of several neutrinos (Waxman & Bahcall 1997). These neutrinos can be detected by the Čerenkov light emitted by high-energy muons produced by their interactions under the surface of the Earth. Guetta et al. (2001) estimated the neutrino production in the framework of the internal shock model and found similar results. They calculated self-consistently the production of synchrotron photons and photo-meson neutrinos from photo-meson interactions of protons with γ-rays within the fireball dissipation region. This calculation takes into account the synchrotron losses of high-energy pions and muons, as well as pair-production by higher energy photons. Since in GRBs the typical photon energy is ~1 MeV, protons with energy of a few ×10^{16} eV are needed to produce neutrinos from pion decay, leading to ~10^{15} eV neutrinos. Their model predicts the detection of ~10 neutrino-induced muons per year over 4π sr correlated with GRBs, in a 1 km^3 detector. The time-averaged neutrino flux predicted in their model is consistent with the flux predicted if GRBs were the source of UHECRs (>10^{19} eV).

Neutrinos with still higher energy (~10^{18} eV) may be produced at a later stage, at the onset of fireball interaction with the surrounding medium. With typical afterglow photon energies of ~0.1–1 keV, neutrinos can reach characteristic energies of ~10^{18} eV if protons can be accelerated to 10^{19}–10^{20} eV. Along with π^+ and π^−, π^0 are produced leading to very-high-energy γ-rays, but as the fireball is optically thick at these energies, the γ-rays probably leak out only at much lower energy, ~10 GeV. These UHE (ultra-high-energy) neutrinos and multiple gigaelectronvolt γ-rays would lag the main GRB by ~10 s and might be detectable. Again, this assumes that protons can be accelerated to ~10^{20} eV in the mildly relativistic reverse shock created when
the expanding fireball collides with the surrounding medium. Vietri (1998a) showed that ultra-high-energy neutrinos (>10^{19} eV) are not only produced during the burst, but also during the afterglow. These neutrinos result from the interactions of UHE protons with photons, which produce pions and then neutrinos. The crucial problem of the losses which can limit the neutrino energy has been considered by Vietri (1998a) and by Rachen and Mészáros (1998; see also Guetta, Spada, & Waxman 2001). Vietri (1998a) considered this issue for external shocks and concluded that neutrinos produced by pion decay are unaffected by losses (adiabatic and synchrotron losses) while neutrinos produced by muon decay in strongly beamed emission are limited in energy but still manage to exceed 10^{19} eV. The conclusion of Vietri is that UHE neutrinos can be produced in afterglows. Rachen and Mészáros (1998) investigated losses (protons, pions, muons) in the internal shock scenario. They found that the cooling of pions and muons in the hadronic cascade imposes the strongest limit on the neutrino energy and that the production of UHECRs and observable neutrino fluxes at the same site can only be realized under extreme conditions. They concluded that GRBs can produce a significant flux of UHE neutrinos but not above 10^{18} eV.

In summary, the production of UHE neutrinos, like the production of UHECRs (sites and processes) is an unresolved issue. UHE neutrinos detected simultaneously with a GRB would provide good evidence that GRBs also produce UHECRs (in the external and internal shock scenarios, the neutrino pulse is essentially simultaneous with the GRB). A very clear and concise review on high-energy neutrinos from GRBs can be found in Waxman (2000) and Mészáros and Razzaque (2006). It treats the acceleration of UHECRs in internal and external (reverse) shocks with constraints on $\varepsilon_B$ and $\Gamma$ to allow the acceleration of protons up to $>10^{20} \text{eV}$. As we have already said these conditions are similar to those deduced from $\gamma$-ray observations: $\Gamma > 100$ and $\varepsilon_B$ close to equipartition. The energy generation rate of UHECRs by GRBs and the production of high-energy neutrinos in internal and external (reverse) shocks and in inelastic $p-n$ collisions are analyzed. Neutrinos with typical energies of $\sim 10^{18} \text{eV}$, $10^{14} \text{eV}$, $10^{10} \text{eV}$ are produced in the afterglow, internal shock and inelastic $p-n$ interactions respectively. Waxman also indicated briefly the implications for GRB neutrino detection. This analysis demonstrated the possibility of a rich production of neutrinos in various energy domains. Even if there has been no detection of neutrino signal from GRBs so far (Achterberg et al. 2008), we can be reasonably optimistic for the near future with the current and future (ice and water) Čerenkov neutrino detectors:

- **Amanda II** the Antarctic Muon and Neutrino Decay Array uses the natural ice at the South Pole as a Čerenkov medium (Stamatikos & Amanda Collaboration 2004). This TeV–PeV muon neutrino detector has been working for 5 years with no indication of point sources of neutrinos. The upper limits are largely above the theoretical predictions (Achterberg et al. 2007a,b, 2008).
- **The large IceCube neutrino detector** has been deployed more than 1 km deep in the Antarctic ice at the South Pole near the existing Amanda detector. This 1 km$^3$ detector is sensitive to all three neutrino flavors and it represents the new generation of neutrino detectors in the tera-, peta-, and even exaelectronvolt...
range (Ahrens et al. 2004); its first-year performance has been described by the IceCube Collaboration (2006).

- Other detectors can be used for neutrino searches: ANTARES a large Čerenkov neutrino detector in the Mediterranean Sea (Aguilar et al. 2006), the RICE experiment using an array of dipole radio receivers (Besson et al. 2007), ANITA (Antarctic Impulsive Transient Antenna), a long-duration balloon payload (Barwick et al. 2006), the Pierre Auger high-energy cosmic-ray observatory (Abraham et al. 2004a), and KM3NeT, a km$^3$-scale detector in the Mediterranean Sea, the successor of ANTARES (Katz 2006).

Despite these large detectors, the chances of detecting ultra-high-energy neutrinos from GRBs are low and detections, if any, will certainly be limited to few nearby bursts. But the detection of these neutrinos which point back to their sources is of major interest because it will bring a better understanding of these sources and of the acceleration mechanisms at work, possibly establishing GRBs as the source of GZK cosmic rays. In Section 9.5 we have raised questions about the nature of the fireball (matter- or Poynting-flux-dominated). The detection of high-energy neutrinos would strongly favor matter-dominated jets. We recommend the review of Mészáros and Razzaque (2006) on the theoretical aspects of high-energy neutrinos associated with GRBs, in order to fully appreciate the richness of this field.

Another source of information which might be crucial in identifying the progenitors of GRBs is the possible detection of a gravitational signal. Like neutrino astronomy this is a speculative venture, but we will see that it is no more out of reach thanks to a new generation of gravitational wave detectors.

9.7 GRAVITATIONAL WAVES

We have seen in Chapter 8, that gravitational waves (GW) might be expected to be associated with GRBs, especially short GRBs believed to result from the coalescence of two compact objects, BH + NS or NS + NS. Regardless of their possible association with GRBs, binary compact object mergers have been studied as potential GW sources (see Mészáros (2006) for a list of references to these works, and Cutler & Thorne (2002) for an overview of GW sources). The detection of a GW signal in coincidence with a GRB would be a very exciting discovery. GWs are emitted from the immediate vicinity of the central engine and they constitute the most direct probes of the progenitor, of its mass distribution and of its nature. In comparison, even the prompt GRB emission originates far from the source ($r > 10^{13}$ cm) and it provides only indirect information about the inner parts of the central engine where the energy is liberated. Major facilities are now in operation for the detection of gravitational waves. The two major interferometers in activity are LIGO (Laser Interferometer Gravitational Wave Observatory; Abramovici et al. 1992, Abbott et al. 2008a) in the US and VIRGO (Variability of Irradiance and Gravity Oscillations; Acernese et al. 2004, 2007) in Europe (Pisa, Italy), but large detectors also exist in other countries,
like GEO 600 in Germany (Strain et al. 2004) and TAMA in Japan (Tatsumi et al. 2007). Coalescing binaries and collapsars with rapidly rotating cores leading to the possible formation of a bar may produce similar GW signals. Davies et al. (2002), van Putten (2001) have indicated that instabilities in the collapsing core or in the accretion disk of a collapsar could lead to stronger GW signals than expected in previous numerical estimates. Davies et al. (2002), for instance, suggested that the core collapse of a massive rotating star may lead to fragmentation of the star with the production of nuclear density lumps. Their coalescence under gravitational radiation gives a BH or NS with a significant kick velocity. With this picture these authors found that core-collapse SNe could be significant sources of GW with signals similar to coalescing NS binaries. Fryer, Holz, and Hughes (2002) found that bar-mode and r-mode instabilities from core-collapse SNe remain among the leading candidate sources for the LIGO II detector. Fryer et al. (2001) and Dimmelmaier et al. (2001, 2002) found nevertheless that the expected luminosity is lower for a collapsar than for the coalescence of two compact objects. It should be noted, however, that a larger uncertainty is attached to the collapsar emission. As an example, Fryer, Woosley, & Heger (2001) noted that their computation of the energy radiated in GW during the collapse of very massive (150–300 $M_\odot$) low-metallicity rotating stars was not very accurate because the BH formation is followed in two dimensions only. In fact, accurate simulations of the development of instabilities and a better estimate of the GW emission would necessitate 3D simulations (Fryer, Woosley, & Heger 2001). 3D calculations of the collapse of rotating stars have been undertaken by Fryer, Holz, and Hughes (2004) and Fryer and Warren (2004). With rotating stars the spherical symmetry is broken and Fryer and Warren (2004) analyzed the consequences of the star rotation. For GWs it leads to signals that are significantly stronger. Fryer, Holz, and Hughes (2004) have revisited their work, taking into account the fact that their first results were obtained for very fast (perhaps too fast) rotating stellar cores. When they consider what they call ‘modern progenitors’ of stellar collapse, they are only marginally unstable to bar instability (if at all) and under these new conditions strong bar-mode instabilities, which are the source of strong GW signals, do not occur. So the bounces of collapsar-driven SNe (GRB-like) would only be detectable when they occur in the Galaxy. They concluded that collapsars are not expected to be strong GW sources given the low rate of such events in the Galaxy, $10^{-5}$ yr$^{-1}$ (Fryer, Holz, & Hughes 2004). This is not far from the conclusions of Kobayashi and Mészáros (2003) who considered both bar-mode instabilities in the disk (around the rotating BH) and the possible effects of asymmetric collapse and breakup. The orders of magnitude of the strain and the frequency of GWs were evaluated for the two possible types of GRB progenitors: compact mergers and collapsars. The possibility of their detection by the advanced LIGO system was estimated (see their Figures 1, 2, and 5 giving the amplitude of the GW signal for 3 of the most promising scenarios: BH–NS, NS–NS, and collapsar). For the collapsars, they concluded that they might be marginally detectable sources of GW with advanced LIGO or VIRGO, in 1 year of observations. These divergent conclusions demonstrate that the possible detection of GW from a massive star collapse is still a very open issue.
Compact binary mergers (BH–NS and NS–NS) are now considered to be the most serious candidates for explaining short GRBs (SHB) (see, however, the reservations of Nakar, Gal-Yam, & Fox (2006) concerning NS–NS systems). The time-integrated GW luminosity has been calculated for merging BH–NS and NS–NS models (Kochanek & Piran 1993, Ruffert & Janka 1998, Nakamura 2000, Kobayashi & Mészáros 2003). The early results of Ruffert and Janka (1998) were not very optimistic. They found that the GW strain would be close to the lower sensitivity limit of the new generation of GW interferometers. Recent rate estimates have been given by van Putten in his book,\(^5\) by Belczynski et al. (2006), and Nakar, Gal-Yam, and Fox (2006). These last authors discussed the consequences of the detection of SHB afterglows for compact binary models and GW detection. The discovery of the first four afterglows of SHBs (see Section 4.11) suggested that these events result from long-lived progenitor systems. They found that the typical progenitor lifetime is long, a few gigayears, and that the local rate of SHBs is larger than 10 Gpc\(^{-3}\) yr\(^{-1}\), perhaps much larger. If progenitors of SHBs are strong sources of GWs this local SHB rate might have direct implications for the detection rate of GW telescopes. Nakar, Gal-Yam, and Fox (2006) evaluated the capability of LIGO, VIRGO, and LIGO II of detecting these sources when the system is made of a NS orbiting a 10 \(M_\odot\) BH (mass ratio larger than 1) and when it includes two NS. Since the first system is the most promising, we will only consider this case below. They found that such systems might be detected as GW sources by LIGO I up to a distance \(\sim 40\) Mpc. The expected detection rate would be \(\sim 3\) yr\(^{-1}\) for the ‘reasonable’ SHB rate we have reported previously. This is a factor of 10 higher than previous upper limits. The expected rate of simultaneous SHB + GW events is a factor of 10 lower: 0.3 yr\(^{-1}\). Of course, this number is strongly dependent on the beaming. With LIGO II, which is more sensitive, the estimates are higher and very encouraging: LIGO II would be able to detect the coalescence of BH–NS systems up to a distance of 650 Mpc. If SHBs originate in such mergers a robust lower limit for the detection of GW is \(\sim 20\) yr\(^{-1}\). For LIGO II, the coincident event rate might be a few per year. Cutler and Thorne (2002) estimated that two simultaneous detections in GW and \(\gamma\)-rays might be expected per year with LIGO II, if BH–NS mergers are responsible for SHBs. Nakar, Gal-Yam, and Fox (2006) noted nevertheless that these estimates may be too optimistic, because the mass ratio between the BH and the NS cannot be too large in order to produce the accretion disk and the SHB (see Miller 2005 and Rosswog 2005). In the case of NS–NS mergers, the rate of coincident GW + SHB events would be dramatically lower: 0.2 event per year. So we have to be cautious about these rates and the hopes they generate.

As these GW signals are weak, we have seen that the detection of an association between a burst of GW and a GRB would be helpful for increasing the strain sensitivity. Kochanek and Piran (1993) noted that the sensitivity is 50% larger in a coincidence experiment compared with a random search. Unfortunately, due to the beaming of GRB emission, we expect few coalescence GW signals accompanied by

---

Some problems encountered by present GRB models

Simultaneous GRB detection. With the same objective of increasing the sensitivity, Finn, Mohanty, and Romano (1999) proposed comparing the correlated output of two GW detectors immediately preceding a GRB to the output at other times. The authors emphasized the fact that this method does not require a priori knowledge of the waveform of the GW burst. This method has already been used to place upper limits on the emission of bursts of GW by GRBs (Astone et al. 2002, see also Abbott et al. 2008b). It is worth noting that even upper limits on the GW strain associated with GRBs would be helpful for constraining GRB models.

After this short analysis of GW signals expected from GRBs, we can say that GW detections are eagerly awaited with the next generation of GW interferometers. Even though the estimates are not completely reliable and the expected signals are low we hope that the near future will have some good surprises in store for us, particularly after the launch of GLAST with its all-sky GRB monitoring, and with the new generations of GW detectors. The coincident detection of GWs and SHBs is a fascinating objective for future missions, with dramatic consequences on our understanding of the nature of GRB progenitors. Since we have only given a short overview of this field, we would like to recommend to the readers interested in a much more complete analysis of these topics two review articles by van Putten: the book that we have already pointed out, and an article in Physics Reports (2005) entitled ‘Gamma ray bursts: LIGO/VIRGO sources of gravitational radiation’.

Conclusions

9.8 SOME PROBLEMS ENCOUNTERED BY PRESENT GRB MODELS

Before concluding with a short analysis of GRB missions which are under development and some discussion about the near future of GRB studies, it is interesting to enlighten these perspectives with a reminder of the most critical open issues which have been reported in previous chapters. This is also a good way to indicate that despite the enormous work (observational and theoretical) which has been done, the development of the popular fireball model has not permitted the mystery of GRBs to be completely solved. The recent results of Swift may even have contributed to rendering our vision of GRBs more opaque.

9.8.1 Origin and structure of the magnetic fields

The origin, the structure, and the strength of magnetic fields in GRB fireball plasmas are important and open questions. They are critical because it is generally accepted that the prompt emission and the afterglow are produced by synchrotron emission which depends strongly on the strength of magnetic fields within the shocked regions. Moreover, in alternative models, which consider Poynting-flux-dominated outflows, the magnetic field is even more crucial as it not only contributes to the synchrotron
emission, but also carries most of the energy. As noted by Piran (2005a), the required magnetic fields may be carried from the inner engine or generated and amplified by the shock.

For the prompt emission, assumed to come from the internal shock region, the energy density is large and the corresponding magnetic fields can also be large: $10^3 - 10^4 \, \text{G}$. These fields can be transported out from the inner engine: the magnetized system consisting of the BH with its accretion disk. The toroidal component of the field decays as $R^{-1}$ and a field of $10^{14} \, \text{G}$ at $10^6 \, \text{cm}$ can be as high as $10^6 \, \text{G}$ at $10^{14} \, \text{cm}$, where the prompt emission occurs (see, e.g., Proga et al. (2003), who show that within the collapsar model MHD effects imply that the toroidal field could dominate over the poloidal field and that the ejecta has to be magnetized). The importance of magnetic fields in the production of GRBs—their role in the gravitational collapse and in the formation of the outflow—has been stressed by Mizuno et al. (2004; see their list of references for these different topics). For magnetically driven outflows with a ratio of magnetic to kinetic energy density of a few to 10%, the field strength is sufficient for efficient production of $\gamma$-rays via synchrotron emission, without the need for local generation of small-scale fields (Spruit, Daigne, & Drenkhahn 2001). Unfortunately, it is not known to what degree the GRB outflow is magnetized. The degree of magnetization is often described with the magnetization parameter $\sigma$ (see for example Zhang & Mészáros 2002a,b, Zhang & Kobayashi 2005). The case $\sigma \to 0$ corresponds to a purely hydrodynamic fireball. This is the situation generally considered for GRBs in which the magnetic field is introduced through the equipartition parameter $\varepsilon_B$. The case $\sigma \to \infty$ corresponds to a Poynting-dominated outflow. In this case the energy dissipation is in the form of magnetic field reconnection and the magnetic fields involved are larger than in a baryonic outflow. Of course, a mix of the two situations can be imagined, which is less demanding for magnetic fields (Spruit, Daigne, & Drenkhahn 2001). In this case a moderate value of $\sigma$ would be the real situation for GRBs, but this is an open question. In the future, the measurement of the polarization of the prompt high-energy emission may provide clues to the nature of the outflow and its degree of magnetization.

In the case of the afterglow, the emission region is located at a distance from the central engine which is much larger than that of the prompt emission. So the expected field of typically $\sim 1 \, \text{G}$ which is needed to explain the emission by synchrotron processes cannot be connected with the central engine. This is explained for instance by Piran (2005a) and Li and Waxman (2006) who showed that if the post-shock magnetic field is inferred to be near equipartition and if the typical field of the interstellar medium is $\sim 1 \, \mu\text{G}$, an amplification of the field energy density (beyond compression) by seven orders of magnitude is needed (Gruzinov 2001). This raises the question of the origin of a significant field in the afterglow since the propagation of the external shock is usually assumed to take place in an environment where no large-scale field is present. Various authors have assumed that the magnetic field has to be locally generated by microscopic processes (Mészáros & Rees 1993, Wijers, Rees, & Mészáros 1997, Thompson & Madau 2000). As collisionless shocks in the afterglow are composed of two flows that cross each other, Medvedev and Loeb (1999) proposed the Weibel instability as the source of the strong magnetic fields in the
afterglow, with an energy density $10^{-5}$ to $10^{-1}$ of equipartition (see also Blandford & Eichler 1987, Gruzinov & Waxman 1999, Wiersma & Achterberg 2004). This instability, which is driven by the anisotropy of the particle distribution function, could also operate in internal shocks, but we have seen that this may not be necessary. It has been argued, however, that even if a strong field is generated in the vicinity of the shock transition (see, e.g., Piran 2005a, Milosavljević & Nakar 2006a,b), there is no evidence that it can persist over the required distance from the shock, about $10^3 \lambda_s$, where $\lambda_s$ is the proton plasma skin depth. In fact it is expected to decay within a few skin depths (Gruzinov 2001). Hence the survival of the magnetic field over large distances is not demonstrated.

Milosavljević and Nakar (2006b) considered the influence of cosmic rays (CRs) on weakly magnetized relativistic collisionless shocks. CRs in SNe can drive turbulence and amplify the magnetic fields upstream of the shock (Bell 2005). The turbulence driven by CRs, on scales much larger than the skin depth, might generate a large-scale magnetic field and produce a magnetized shock. The authors considered that the origin of the magnetic fields involved in the synchrotron emission of the afterglow in the external shock model is plausibly in the CR-driven turbulence. Li and Waxman (2006, and references therein) also discussed the question of the growth of the field length scale and they studied the constraints on the upstream field strength imposed by X-ray afterglow observations: $B \gg 0.05$ mG for a density $n = 1$ cm$^{-3}$ (upstream plasma density) up to 0.2 mG in several cases. These results suggest that either the shock propagates into a highly magnetized fast wind or that the pre-shock magnetic field is strongly amplified, most likely by the streaming of high-energy shock-accelerated particles (Bell 2004). The first solution is not obvious because the WR involved in the collapsar model are not expected to have magnetically driven winds. Moreover, we have seen that in general afterglow observations do not indicate propagation of the ejecta into a wind, but rather into a uniform density medium. Hence the solution might be the amplification of the field ahead of the shock by the streaming of high-energy particles as analyzed by Bell (2004).

The generation of magnetic fields in collisionless shocks is thus an important issue, and we have indicated that the Weibel instability may play a central role in this context, but complete analyses show that we are still far from a robust solution. At this time (2008), the degree of magnetization of the outflow and the mechanisms of its magnetization remain open issues. It has been suggested (Sagiv, Waxman, & Loeb 2004) that spectro-polarimetric observations in radio and IR bands during the early stages of the afterglow might be used as unique probes of the structure and strength of the magnetic field (see also Nakar & Piran 2004).

### 9.8.2 Particle acceleration and the spectral index of the electron energy distribution

Another question concerns the acceleration of particles in collisionless shocks and the value of $\rho$, the spectral index of the electron energy distribution. In fact, the microphysics of these shocks is a major open issue in GRB physics with serious implications for the means of accelerating particles and of generating and sustaining magnetic fields (see the review of Waxman 2006a). Even the necessity of shocks to
explain prompt GRB emissions has been questioned recently by Kumar et al. 2007 (see Section 9.8.7), so let us comment here on this central question of shocks for particle acceleration. Waxman (2006a) noted that a good understanding of particle acceleration in relativistic shocks exists for the test particle approximation, but a theory that self-consistently describes the acceleration of the particles and the generation of electromagnetic waves is still missing; the accelerated particles probably have an important role in generating and maintaining the inferred magnetic fields. The efficiency of these shocks and even the presence of Fermi acceleration at GRB shocks have been questioned. From their simulations of first-order Fermi acceleration processes in ultra-relativistic shock waves, Niemiec and Ostrowski (2006) found that the spectra obtained for the ‘realistic’ conditions assumed in their simulations do not converge to a ‘universal’ spectral index. They suggested that the role of first-order Fermi acceleration in sources hosting relativistic shocks must be reassessed. In particular, they believe that this mechanism cannot be the main one which accelerates the electrons radiating in GRB afterglows. Moreover Nishikawa et al. (2003) concluded that most of the electron acceleration takes place behind the jet front and cannot be characterized as Fermi acceleration. Nevertheless they showed that some Fermi acceleration may occur at the jet front.

Thus it is not surprising that other acceleration mechanisms have been proposed. Nishikawa et al. (2003) and Hededal et al. (2004), for instance, have studied an acceleration mechanism which is different from Fermi acceleration. Both consider the Weibel two-stream instability operating in relativistic collisionless shocks. Hededal et al. (2004) noted, however, that this mechanism has not been conclusively demonstrated to occur in ab initio particle simulations (see also Baring & Braby 2004). The role of this instability has also been disputed by Milosavljević and Nakar (2006a) who considered that CRs drive turbulence into the region upstream of the shock on scales much larger than the skin depth and that this turbulence generates a large-scale magnetic field that quenches the transverse Weibel instability. These authors think that the picture involving these instabilities requires revision. So the situation is complex.

Additional questions connected with electron acceleration are the fraction of the electron population which is accelerated and the value of the spectral index $p$, which characterizes the power-law spectrum of accelerated electrons. From the X-ray spectral slope of 15 afterglows in the pre-Swift era, de Pasquale et al. (2006) found a value $p = 2.4 \pm 0.2$. On the other hand, Panaitescu et al. (2006b) found indices $p$ ranging from 1.3 to 2.8 for nine Swift afterglows, and a single electron index $p = 2.1 \pm 0.1$ can only be obtained if four bursts are removed. Panaitescu et al. (2006b) thus suggested that shock accelerated electrons do not have a universal energy distribution. These authors also noted that this appears to be in agreement with the wide dispersion of the high-energy spectral slopes of BATSE bursts ($\Delta \beta \sim 2$, where $\beta = p/2$; see Chapter 5) and with the wide range of optical post-break decay indices of BeppoSAX afterglows ($\Delta \alpha = 1.6$, where $\alpha = p$; see Chapter 5). Shen, Kumar, and Robinson (2006) arrived at the same conclusion, using the high-energy photon spectra of BATSE GRBs. The comparison of the high-energy spectral indices of time-resolved spectra and of time-averaged spectra shows that time averaging does
not contribute to the observed dispersion, implying that the distribution of \( p \) is intrinsically broad. The distribution of \( p \) during the X-ray afterglows is also broad. Finally the comparison with the distribution of \( p \) for 44 blazars leads to the same conclusion, namely that the \( p \) distribution is broad and not consistent with a single universal value (Shen, Kumar, & Robinson 2006). But, as indicated in their conclusions, there are various reasons to have a value of \( p \) that is not universal. Moreover, we have seen that other acceleration mechanisms have been proposed, for instance the Weibel two-stream instability mentioned above, but it is not known whether a single universal value of \( p \) is predicted by this mechanism.

We must conclude that the acceleration mechanism at work in relativistic collisionless shocks and the magnetic field generation are not fully understood, leaving uncertainties in the way the electrons are accelerated, in the maximum energy they can reach, and in their spectral index. One consequence of these uncertainties is that it is not obvious whether GRBs can accelerate protons to ultra-high energies, and produce UHECRs and high-energy neutrinos.

### 9.8.3 The microphysics parameters

Another important point of the fireball model is the microphysical parameters \( \varepsilon_e \) and \( \varepsilon_B \), which represent respectively the ratio of the kinetic energy density of the electrons and of the magnetic energy density to the total energy density dissipated behind the shock. The values of these parameters are broadly distributed over at least an order of magnitude but the typical values are \( \varepsilon_e \sim 0.2 \) and \( \varepsilon_B \sim 0.001 \) (see for instance Panaitescu & Kumar 2001, Yost et al. 2003). We have already mentioned the Weibel instability to accelerate the electrons in the shock. Numerical simulations of this mechanism establish a value of \( \varepsilon_B \) at the level of few times \( 10^{-3} \) to \( 10^{-4} \) (Medvedev 2006), in agreement with the reported values.

But what is the relation, if any, between \( \varepsilon_e \) and \( \varepsilon_B \), and are these parameters constant for a given GRB? To the first question Medvedev (2006), studying the electron acceleration in relativistic shocks, gave an answer through his relation \( \varepsilon_e \propto \lambda \varepsilon_B^{1/2} \) where \( \lambda \) is constant, \( \sim 1 \), in relativistic baryon-dominated shocks. Using the GRBs analyzed by Panaitescu et al. (2006b) he plotted \( \varepsilon_e/(\varepsilon_B)^{1/2} \) for these bursts, taking for each of the 10 bursts the best-fit afterglow model, which is not the same for all the bursts. This ratio clusters around \( \varepsilon_e/(\varepsilon_B)^{1/2} = 1 \) (Figure 9.18). So, for a typical value \( \varepsilon_B \sim 10^{-3} \) the corresponding value of \( \varepsilon_e \) would be \( \varepsilon_e \sim 0.03 \). To the question of the relation between \( \varepsilon_e \) and \( \varepsilon_B \), an answer can thus be given if magnetic fields are generated at the shock by Weibel instabilities. The physical reason behind this relation is that the current filaments formed by protons produce at the same time magnetic fields and electrostatic fields which accelerate electrons. Of course the sample is too small to be conclusive and we have seen the objections to Weibel instabilities as a way of accelerating electrons in the shocks.

Concerning the second question (the constancy of \( \varepsilon_e \) and \( \varepsilon_B \)), we have already indicated that some authors have introduced a dependence of \( \varepsilon_e \) and \( \varepsilon_B \) with time, or with the Lorentz factor \( \Gamma \), to explain the shallow decay phase observed in the early X-ray afterglows of many GRBs, by the Swift XRT (see Section 7.2.4). Panaitescu et
Figure 9.18. The ratio $\varepsilon_e/\sqrt{\varepsilon_B}$ for 10 GRBs analyzed by Panaitescu (2006b). The figure gives the ratio for the best-fit model (with the smallest reduced $\chi^2$ value) for each burst. As $\varepsilon_B$ varies by two orders of magnitude, $\varepsilon_e$ varies by only one order of magnitude. Clearly the data corresponding to this ratio are clustered around unity with little scatter. For Medvedev (2006) the quality of this correlation favors the Weibel shock theory, which explains the origin of magnetic fields in relativistic shocks by current filaments formed by protons moving roughly at the speed of light. These filaments also produce electrostatic fields which can accelerate electrons.

al. (2006a) noted that the introduction of a sudden change in the evolution of microphysical parameters to explain the different behavior of X-ray and optical afterglow light-curves seems ad hoc because the evolution of the microphysics parameters is chosen to reproduce the light-curves but it has no physical basis. They considered another scenario to be more realistic: the microphysical parameters might have a steady evolution with the Lorentz factor of the blastwave. In that case, the break in the X-ray light-curve arises from the cessation of energy injection into the forward shock (see for example Nousek et al. 2006, Panaitescu et al. 2006b, Zhang et al. 2006). The lack of an optical break at the time of the X-ray break determines the dependence of the microphysical parameters with the Lorentz factor (see Section 7.5). Another study which was used to constrain the intrinsic and environmental parameters was done by Yost et al. (2003) for four GRBs. They noted a striking diversity in the fitted microphysical parameter values. In their analysis, for instance, the magnetic energy fraction varied from 0.2% to 25%. Non-universal microphysics parameters thus seem to be the rule. Moreover, they tried to determine the effect of a dependence of $\varepsilon_B$ on $\Gamma$, the bulk Lorentz factor, $\varepsilon_B \propto \Gamma^\alpha$. They found a considerable flexibility in the values of $\alpha$, with $\alpha$ in the range $[-2; 1]$.

The microphysical parameters, $\varepsilon_e$ and $\varepsilon_B$, their values and their possible evolution within a given burst are at the center of the fireball model. The detailed studies of
the early afterglow phase with Swift have revealed unexpected complexity and raised new questions concerning this way of modeling the properties of the ejecta.

9.8.4 $\Gamma$, the bulk Lorentz factor of the fireball

$\Gamma$, the bulk Lorentz factor of the fireball, has a value which is generally considered to be between 100 and several hundred. Using the onset of two GRBs observed with the robotic telescope REM, GRB 060418 and GRB 060607A, Molinari et al. (2007) claimed to make a direct measurement of the initial Lorentz factor of the fireball: $\Gamma_0 \sim 400$ for the two bursts. For this calculation, they assumed typical values of the circum-burst density ($n = 1 \, \text{cm}^{-3}$) and of the radiative efficiency ($\eta = 0.2$); the result was only weakly dependent on these values. These observations also allowed an estimate of the baryonic mass of the outflow, $\Delta M = 7 \times 10^{-4} \, M_\odot$ and the deceleration radius $R_{\text{dec}} \sim 10^{17}$ cm. These values of $\Gamma_0$ are in agreement with lower limits given for instance by Lithwick and Sari (2001) or Zhang et al. (2006). Lithwick and Sari (2001) computed lower limits on $\Gamma_0$ with the requirement that GRBs must be optically thin. They considered different constraints from $\gamma - \gamma$ annihilation and from Compton scattering of photons by the $e^+/e^-$ created by annihilation of photon pairs. From these constraints, their Table 3 gives the lower limits on $\Gamma_0$ for 13 GRBs with high-energy photons (detected by Egret) and/or redshifts. These lower limits range approximately from 100 to 400. Similar constraints have been derived by Woods and Loeb (1995), and Baring and Harding (1997) who obtained $\langle \Gamma_{\text{min}} \rangle \sim 100$, assuming that GRBs have to be optically thin for photon–photon pair production at all observed energies. More recently Zhang et al. (2006) found $\Gamma_0$ to be between 100 to 200($\eta\gamma$)$^{-1/8}$.

GLAST, working in the energy band 30 MeV–300 GeV, will be able to explore the high-energy photon domain with better sensitivity than Egret. We can expect to see cut-offs in the high-energy photon spectrum. Baring (2006) analyzed this question, trying to identify how the attenuation of photons by pair creation internal to the source generates distinctive signatures. He concluded that the energy of spectral breaks and the associated spectral indices will provide information to constrain the bulk Lorentz factor $\Gamma$ of the GRB at a given time. In this search not only GLAST, but also ground-based Čerenkov telescopes, will be of prime importance. If there is no big surprise, we can say that the determination of $\Gamma$ seems now just a question of time, with sensitive high-energy $\gamma$-ray instruments coming into operation in 2008.

9.8.5 The density of the circum-burst medium

The density in the circum-burst environment represents another puzzle. Because the collapsar model, which is successful for explaining long GRBs, involves massive star explosions, one naturally expects an environment with a stellar wind and a density $n(r)$ varying with $1/r^2$ dependence. While this is observed for some bursts, at least half of them seem to occur in surroundings with uniform density profiles (Fryer, Rockefeller, & Young 2006). To explain this result these authors have modified the standard WR mass loss paradigm, but they could not produce constant density
profiles. Hence they proposed, without abandoning the collapsar model, to focus on the He merger model (see their Figures 5, 6, and 7 for different scenarios), which, in their opinion, provides a solution to this problem. But there is not a strong consensus on a uniform density profile. For instance Zeh, Klose, and Kann (2006) concluded from an analysis of 16 afterglows with well-sampled light-curves that in most if not all cases the data are in agreement with a wind model.

Another question, examined by Scalo and Wheeler (2001) is the large variety of ambient densities, from 0.001 to 30cm\(^{-3}\) or more. The lowest densities seem incompatible with bursts within or near molecular clouds or with dense stellar winds. The densities expected in these environments would have to reach values as large as \(10^4 - 10^6 \text{ cm}^{-3}\), which are very far from the lowest densities which have been reported: \(10^{-1} \text{ to } 10^{-3} \text{ cm}^{-3}\). To avoid this inconsistency Scalo and Wheeler (2001) argued that low-ambient-density regions naturally exist in areas of active star formation, for instance in the interior of super-bubbles, and they proposed that GRBs occur inside pre-existing super-bubbles. In this way they explained the very large range (four to five orders of magnitude) of ambient densities and their lower and higher values. They argued that low ambient densities for some afterglows do not preclude a priori massive star progenitors for GRBs. Nevertheless, the contamination of the low-density environment (of the super-bubble) by the stellar wind from the massive star remains a severe problem and the lack of any evidence of such a wind implies very low wind densities. Finally, there is no simple and definite explanation of these densities in the context of current models involving massive stars exploding in dense regions, which are expected to have strong stellar winds. These are very complex issues, which involve all the characteristics of the progenitors and their environment; for instance, the low metallicity of massive stars might modify dramatically the problem of mass ejection before the final phases of their evolution.

More generally, the properties of GRB environments are poorly known and many questions remain, for instance: Is the ambient medium clumpy? Is it magnetized, and to what degree?

Here again there is no trivial solution to reconcile the properties of the GRB environment with our understanding of the progenitors of long GRBs. Now we will turn to other problems in relation to the results of Swift on the early GRB afterglow. We have analyzed these questions in Chapter 7; let us recall the main issues below.

9.8.6 The early X-ray afterglow: problems posed by the fireball model

The shallow X-ray phase

The shallow phase (see Section 7.2) has been explained by late energy injection. We have shown that late energy injection affects the radiative efficiency, \(\epsilon_\gamma\), during the prompt phase. In the absence of late energy injection, the radiative efficiency is high but still compatible with the internal shock model, whereas with late energy injection it might reach 0.9, a value which is clearly not compatible with the internal shock model. This shallow phase and its classical explanation by a late energy injection can thus be in conflict with the popular internal shock model. But, as we have discussed, the conclusions are not clear-cut. For instance Fan and Piran (2006) concluded that the efficiency of converting the kinetic energy into \(\gamma\)-rays is moderate and does not
challenge the standard internal shock model ($\varepsilon_{\gamma} \sim 0.5$, not 0.7–0.9), even when late energy injection is considered to explain the shallow phase. On the contrary, very high values ($\varepsilon_{\gamma} \sim 0.7–0.9$) have been given for instance by Ioka et al. (2006), which are unreasonably high for the internal-shock model. The authors discussed this issue and proposed solutions to avoid energy injection: one of them is to have $\varepsilon_{\gamma}$ increasing during the shallow decay, $\varepsilon_{\gamma} \sim t^{1/2}$, combined with the decay of magnetic field, $\varepsilon_{B} \sim t^{-0.6}$, so that the absence of an optical break at the end of the shallow X-ray decay suggests a time-dependence of the microphysics parameters preferentially over other solutions (see Section 7.5). A consequence would be the suppression of the optical emission from the reverse shock (see Roming et al. 2006, reporting that the early optical afterglow is not detected in a large number of cases, in contrast to the bright but very rare optical flashes discovered in GRB 990123 and GRB 080319B). The questions raised by this shallow phase are so important that various authors have proposed to completely change the afterglow paradigm. Genet, Daigne, and Mochkovitch (2007) and Ulm & Beloborodov (2007) explained the entire X-ray afterglow by a long-lived reverse shock. Ghisellini et al. (2007) proposed a late-prompt emission (due to late internal dissipation) adding up with the standard X-ray afterglow. Panaitescu (2008) suggested that the forward shock photons could be upscattered to X-ray energies by a relativistic outflow located behind the leading blastwave (IC and bulk upscattering).

The jet breaks

Another puzzling question is the nature of the temporal breaks observed in optical and X-ray afterglows. Before Swift, the breaks seen in the optical band were considered as achromatic, proving the presence of jet. The time of the break provided a convenient way of measuring the opening angle of the jet. With Swift, only a few breaks are consistently with being achromatic jet breaks. We have discussed this crucial point in Section 7.4. Often X-ray breaks are not observed after the end of the shallow phase, and many bursts show a single power-law decay lasting for 10 to 70 days. Are the X-ray breaks masked by some additional source of X-ray emission or delayed up to several days where the follow-up is scarce?

This question of the nature of the temporal breaks and the absence of clear achromatic breaks has consequences for the jet model. Is there a jet? If the jet exists how can we detect it and determine its opening angle and its structure?

This problem of jets and their possible detection in the afterglow was eagerly disputed even before the results of Swift (e.g. Moderski, Sikora, & Bulik 2000, Dado, Dar, & De Rújula 2002). These authors concluded that the breaks in afterglow light-curves are weaker and smoother than the break analytically predicted, even if the possibility of having more prominent breaks exists. So it is quite possible that, even if GRB jets exist, their impact on afterglow light-curves is too shallow to allow the determination of the jet break time. Given this difficult situation, it is perhaps interesting to look at the orphan afterglows which are considered to be a way to measure the average opening angle of the jets. But here again the results are not really illuminating, in the optical band or in the radio band. Soderberg et al. (2006), for instance, used late-time radio observations of 68 local SNe of type Ibc. None of these objects exhibited radio emission attributable to off-axis GRB jets. This allowed them
to constrain the beaming factor, $f_b^{-1} < 10^4$, corresponding to an average jet opening angle $\theta_j > 0.8^\circ$. Levinson et al. (2002) and Gal-Yam et al. (2006), on the other hand, provided a lower limit based on a deep search for radio afterglows, $f_b^{-1} > 60$. Dalal, Griest, and Pruet (2002) discussed the difficulty and perhaps the impossibility of seeing orphan optical afterglows because, at late times the beaming angle scales with the jet angle ($\Omega_{\text{opt}} \propto \Omega_{\text{jet}}$). Therefore the number of expected afterglow detections is the same for moderate jet angles ($20^\circ$) and for very small jet angles ($\ll 0.1^\circ$), and the ratio of detected orphan afterglows to GRBs cannot be used to measure $\theta_{\text{jet}}$. Moreover, the authors showed that detectable orphan afterglows are mostly contained in a cone of opening angle not much larger than $\theta_{\text{jet}}$ (due to the depth of optical searches which go down to $R = 23$ at best). Under these conditions, one cannot expect many more orphan afterglows than on-axis afterglows. Dalal, Griest, and Pruet (2002) concluded that radio searches may give the best hope of finding the missing orphans.

These examples indicate that, from the Swift observations, we have no real clue about the shape and size of GRB jets, despite the consensus reached in the BeppoSAX era, when many afterglows seemed to follow the predictions of the simplest version of the external shock model.

**Activity of the central engine**

Late X-ray flares provide convincing evidence of the continuing activity of the central engine and of the ejection of material with a broad range of Lorentz factors. This is another important Swift discovery. Since the models of the central engine are less developed than the models of the afterglow, it is too early to say whether this prolonged activity of the central engine will require significant modification to the standard model and will change our understanding of the progenitors.

### 9.8.7 The progenitors

Another consequence of Swift discoveries relates to the possible progenitors of short and long GRBs. We have seen that the shallow phase and the X-ray flares require a long-lived central engine, with activity extending for several hours. We have also shown in Section 7.4 that this is not easy to explain. Another way would be to have short-duration energy injection but with an extended range of Lorentz factors for the outflow (see for instance Zhang & Mészáros 2002b). This question of continuous or sporadic energy injection is an issue for the long GRB progenitors as well as the short ones because the short bursts also produce X-ray flares. This is perhaps not surprising since both cases involve a black hole (BH) with a transient torus of matter, the difference being the mass of the black hole and its environment at the time of the GRB trigger. There are still fundamental questions about the progenitors of long and short GRBs.

**Long GRBs**

The collapsar model has been well studied and confirmed by the undisputable association of several nearby GRBs with type Ic SNe, proving the presence of a
massive star (WR-like) explosion and suggesting that all (or nearly all) long GRBs have an underlying type Ic SN. On the other hand, it seems established that only a small fraction of type Ic SNe can produce long GRBs. One difference between GRB and non-GRB supernovae could be the very high amount of angular momentum needed to launch a GRB. The strong stellar winds usually associated with WR stars decrease the angular momentum of the star, whereas in the collapsar model a rapid rotation is needed to achieve the formation of a BH with a torus of matter. This is required to allow the matter to stay in orbit around the BH for many seconds, instead of falling into it in milliseconds. This rapid rotation has dramatic consequences for the type of source, binary or single stars, as explained in Section 8.3.

For single stars the question of the rapid rotation is not trivial. Solutions might be found by invoking low-metallicity massive stars, which have lower mass-loss rate. Maeder and Meynet (2001) for instance have shown that massive stars (between 9 and 60 \( M_\odot \)) with low \( Z \) (0.004 \( Z_\odot \)) maintain very high angular momentum (close to break-up) till the end of their life. But the hydrogen envelope has to be liberated because it will be difficult for the jet to pierce the stellar envelope if too much matter remains around the collapsar. A possible solution could be an intense convection, allowing hydrogen to be burned into helium. This would eliminate the necessity of a wind which unavoidably takes away some angular momentum. Moreover, some anisotropy of the stellar wind may be envisaged; if the mass loss is predominantly at the poles, smaller angular momentum losses will result. Unfortunately there is also the problem of magnetic torques, which can brake the core. Petrovic et al. (2005), for instance, raised the critical problem of the role of magnetic fields, emphasizing that their introduction in the models might have significant impact. This was also the conclusion of Heger, Woosley, and Spruit (2005), namely that single stars cannot make GRBs because of magnetic torques, which decrease the final rotation rate by a large factor (30 to 50) when compared with the non-magnetic case. Another problem reported by Fryer et al. (2007) is the necessity for the progenitor to lose not only its hydrogen envelope but most of its helium envelope as well. Hence many problems exist for single-star progenitors and we have given a few examples that illustrate them.

Binary systems might be the solution, especially if magnetic effects are as severe as predicted. Van Dyk, Hamuy, and Filippenko (1996) had already suggested by 1996 that type Ibc supernovae were formed from relatively close binaries rather than single stars, since the interactions with a companion star can severely deplete the hydrogen envelope prior to explosion. Specifically, a binary system may allow removing the hydrogen envelope without losing too much angular momentum and could even inject additional angular momentum (Fryer & Woosley 1998). This was also the conclusion of Podsiadlowski et al. (2004) who studied how the presence of a companion affects the final core structure of a massive star, with consequences for the subsequent SN explosion. They noted that binarity affects many properties of the stellar core, like its rotation rate, the size of the helium core or the mass of the iron core.

Other questions remain and without being exhaustive we can indicate, following Woosley and Bloom (2006):
The role of metallicity, which is connected with the conservation of the angular momentum. We have seen that strong winds in the WR phase may remove too much angular momentum: low-metallicity stars provide a way of avoiding these strong winds, which are driven by the radiation pressure of UV photons on metals in the stellar atmosphere (for the connection between wind mass-loss rates and stellar metallicities see, for instance, Nugis & Lamers (2000) or Vink & de Koter (2005)).

The duration of the central engine activity: Is the emitted power continuous or episodic at late time? The fact that matter ejection can be incomplete in the SN might lead to late phases of accretion and energy injection.

The contribution, if any, of rapidly rotating (\(P \sim 1\) ms) pulsars with very high magnetic fields. These models would be naturally associated with highly energetic Poynting flux outflows.

Without returning to the discussions of Chapter 8 it is clear that even for the collapsar model, which seems to be a very good candidate to explain long GRBs, many questions are still awaiting answers. This is not so surprising when we consider how difficult it is to understand the most common varieties of supernovae. Woosley and Bloom (2006) considered that rotation and magnetic fields may represent the additional physics which has to be introduced to understand SN explosions and GRB progenitors as well.

**Short GRBs**

Compared with long GRBs, the situation of the potential progenitors of short GRBs is not well clarified, even though BH + NS and NS + NS coalescences are the favorite models. Today it cannot be excluded that short GRBs might have another origin besides compact binary mergers, and alternative models have been proposed: accretion-induced collapse of a rapidly rotating NS into a BH (MacFadyen, Ramirez-Ruiz, & Zhang 2005), magnetar, or quark star (e.g. Nakar 2007).

Coalescences involving NS and BH remain, however, the best-studied models. Simulations of the last moments of such systems are becoming more and more sophisticated, with full general relativity treatments and realistic equations of state, and these possible candidates are becoming quite convincing. Nevertheless the observations, particularly the most recent ones obtained with Swift, are puzzling. The few afterglows of short bursts which have been obtained show that X-ray flares are observed in short bursts, as well as in the long GRBs. Moreover, some short bursts have a long tail at low energy, lasting tens of seconds. At least for these two properties, the distinction between these two types of events seems difficult to justify. On the other hand, different types of galaxies (elliptical, star-forming, metal-poor late dwarfs) can harbor short bursts, whereas today long GRBs appear to occur only in star-forming regions of actively star-forming galaxies. Short bursts also seem to be more diffusely positioned around their host galaxies than the long GRBs. The models developed by Belczynski et al. (2006) suggest that the population of short GRB progenitors is diverse in terms of merger times and locations, but the existing sample is too small to be certain of these conclusions. The same difficulty is present in the
redshift determination. The first redshifts measured for SHBs were much smaller than those of long GRBs, varying from $z = 0.17$ to $z = 0.54$, compared to $\langle z \rangle = 2.8$ for long GRBs. But now that more data have been accumulated higher redshifts are obtained for SHBs.

These observations, combined with the diversity of possible progenitors (see Chapter 8), make the overall situation rather complex. A major concern for merger models is the need to have a central engine active for a duration significantly longer than the duration of the initial prompt $\gamma$-ray flash, the natural timescale for the activity of the central engine being typically $\lesssim 1$ s. In connection with this problem, Lee and Ramirez-Ruiz (2007) noted that magnetically powered events may offer a longer staying power. Their review of the nature of the progenitors and of the processes leading to the formation of the central engine is highly recommended. It gives the most complete view of short GRBs and their progenitors, following the discovery of the first afterglows. The discussion on the progenitors of short GRBs can also be extended and enlightened by the review of Nakar (2007). Progress in this field will certainly come from detailed observations of a large sample of nearby short GRBs, with the determination of the type of their hosts and the study of the immediate surroundings of the burst. As discussed in Section 9.7, the observation of a gravitational signal coincident in time and direction with a short GRB would be invaluable.

To close this long, but not exhaustive, list of questions, it is worth mentioning that the nature of the progenitors of short and long GRBs would strongly benefit from complementary observations and advanced theoretical modeling. Observationally, high-energy neutrinos and gravitational waves coincident with GRBs might contribute to provide deeper insight into the nature and activity of the central engine, since they are emitted much closer to the heart of the progenitor than the GRB itself. Theoretically, we need improved simulations of the collapse of massive stars and of the coalescence of compact stars, with realistic equations of state, neutrino transport, 3D calculations, and full consideration of general relativistic effects, and of the role of the magnetic field.

9.8.8 The standard fireball model

Finally, it can be said that the most puzzling questions have been raised by the detailed observations of the early X-ray afterglows of long GRBs performed by Swift. Their impact on the popular fireball model has been very important and may be sufficient to cast some doubts on it. This is perhaps the reason why new and different models have been proposed by Ghisellini et al. (2007), Panaitescu (2008), Genet, Daigne, and Mochkovitch (2007) and Uhm & Beloborodov (2007).

More dramatic conclusions about the standard fireball model itself, with internal shocks explaining the prompt $\gamma$-ray emission, have been raised by Kumar et al. (2007). They studied the nature of the outflow in 10 long GRBs with durations between 11 and 126 s and concluded that the prompt GRB emission cannot be explained by synchrotron emission in a shock-heated medium, nor by the synchrotron self-Compton process in relativistic shocks. For these authors the prompt $\gamma$-ray emission cannot be produced in internal or external shocks; they concluded that the production of $\gamma$-rays by mechanisms involving shock physics has problems. Hence
they suggested processes which do not require shocks, an attractive one being the transport of energy in GRB outflows by magnetic fields and the emission of radiation following the dissipation of this energy farther from the source. The Poynting-flux model has many advantages. As analyzed quite well by Spruit and Drenkhahn (2004), powering GRBs with a magnetic energy flux is an attractive scenario since magnetic fields can carry energy across a vacuum with an extremely low amount of baryon contamination. In addition, magnetic energy dissipation can lead to a prompt non-thermal emission with an efficiency of 50%. The dissipation process can also accelerate the flow to the observed Lorentz factors. Moreover, the high magnetic field strength in the dissipation region is quite favorable for efficient synchrotron emission.

As an example the electromagnetic model of Lyutikov and Blandford (2003; see Chapter 5) makes it possible to explain the generation of γ-rays at a distance of typically $3 \times 10^{16} \text{ cm}$ from the center of the explosion (cf. $10^{15} - 10^{16} \text{ cm}$ found by Kumar et al. (2007) in their analysis of 10 GRBs). If these results are confirmed by a significantly larger sample, this would be a dramatic conclusion for the standard model, which explains the prompt emission by internal or external shocks. Spruit and Drenkhahn (2004) concluded that magnetic powering of GRBs is likely to become an important development in GRB modeling. This conclusion seemed also shared by Zhang (2005) who wrote that we seem to be collecting evidence that at least some GRBs are strongly magnetized or even Poynting-flux-dominated. This reflection is supported by the paucity of GRBs detected by UVOT at early times, suggesting that the reverse shock emission is strongly suppressed.

The standard model must also face other problems, like the apparent high gamma-ray efficiency for some bursts or the flat injection phase in early X-ray afterglows. In defense of the standard model it is very important to say here that it is the only model which has been tested and simulated extensively, and it is also the one which has the largest predictive power, by far. In any case more data, which are being accumulated by Swift, are needed to clarify these important issues. With nearly 400 GRBs detected by Swift, it is important now to perform detailed statistical studies of this large GRB population (see for example O’Brien et al. 2006, Chincarini et al. 2007). Such studies are invaluable for determining the general properties of GRBs, the characteristics of their environments and of their host galaxies, such as redshift, type, color, mass, and metallicity, in order to better understand the progenitors and the GRB emission mechanisms which may become central to theorist’s concerns. Such studies will allow us to make the best use of the large GRB sample, which is being collected year after year by Swift. In parallel, efforts to maintain active and early follow-up campaigns are strongly recommended.

### 9.9 PERSPECTIVES

Wishing long life to Swift with undiminished performance, the launch of GLAST/Fermi in 2008 offers a unique coverage of the γ-ray spectrum up to energies of a hundred gigaelectronvolts and a capability of getting accurate determinations of $E_p$, which appears to be a very important parameter. The progressive start of a new generation of ground-based Čerenkov telescopes capable of measuring very high-energy photons (beyond the range of GLAST), of high-energy neutrino detectors and
UHECR observatories, represent very useful new tools. The association of these facilities with a new generation of interferometers for GW detection will allow exceptional and never before achieved searches for all the possible emissions of GRBs. We can thus expect that major discoveries will enlighten the field of GRB studies, which has already benefitted so much from the prestigious past missions CGRO, BeppoSAX, HETE, and now Swift.

What does the future hold? On what specific points should the next missions concentrate? We tentatively identify below three directions for the post-Swift and GLAST/Fermi period. Before discussing these perspectives, we would like to point out one present difficulty of GRB studies. With Swift, GRBs are detected at a sufficiently high rate to bring to light the problem of follow-up campaigns. This problem can be illustrated with two questions: the existence of achromatic breaks and the nature of the ejecta. The excellent X-ray follow-up provided by Swift–XRT contrasts with the poor optical follow-up available for the majority of GRBs. Few GRBs benefit from high-quality X-ray and optical follow-up, extending over several days in more than one optical filter. Such data are crucially needed for the unambiguous identification of the achromatic jet breaks predicted by commonly accepted GRB models. Concerning the nature of the ejecta, Nakar and Piran (2005) have proposed combined measurements of the early optical afterglow and radio flares to constrain the physical parameters of the original ejecta and the microphysical parameters of the region emitting the reverse shock. This requires a quick response from ground-based telescopes in the optical and radio bands. The authors believe that this would be very useful for clarifying a central question: the nature of the relativistic outflow—baryonic or Poynting flux? Again, such observations are only available for a few GRBs.

After these slightly pessimistic words let us be more optimistic and discuss some observational perspectives for the near future. Thanks to the launch of Swift and the time that has passed since then, and without returning to GLAST/Fermi with its new scientific objectives (Zhang 2007a), we propose emphasizing three domains which deserve deeper investigation:

1. Increasing the number of well-localized short GRBs accompanied by fast and complete follow-up multi-wavelength campaigns. The fraction of short GRBs observed by Swift is at present only 10%, while it was around 30% for BATSE (this difference can be explained by the fact that coded mask localizations—for Swift—require photon fluences which are larger than those needed for the simple detection of an excess with BATSE). We believe that short GRBs are now quite exciting, after the discovery of their afterglows. They can be studied with the same advantages which until recently were reserved for long GRBs, and we certainly have a lot to learn from their detailed study.

2. A second direction would be to concentrate the efforts on highly redshifted GRBs. We have seen that up to 10% of Swift GRBs could be at high redshift ($z > 5$). To be sure of detecting the afterglows of these bursts it would be very desirable having follow-up observations in the NIR: photometry and fine spectroscopy. This would allow defining the redshift and through NIR spectroscopy determining the absorption lines in the afterglow and the emission lines of the host galaxy (both for the same burst, which is very rare today). Hence regions of
early star formation and the environment of the burst in the host galaxy would be accessible, allowing studies to be carried out on the first generations of stars and the evolution of metallicity associated with primordial stars in the early Universe. These objectives may have been achieved already with Swift for a limited number of GRBs. The main effort required is to make large optical, NIR, and radio telescopes available for some selected GRBs which present the typical properties of highly redshifted bursts.

(3) A significant effort is also needed to achieve sensitive polarimetric studies during the prompt and early afterglow phases. For instance, Sagiv, Waxman, & Loeb (2004) argued that polarimetric observations during the early afterglow (where plasma effects on the polarization of propagating synchrotron radiation are significant), particularly of the reverse-shock emission, may place strong constraints on the structure and strength of the magnetic field within the fireball plasma. Polarization may also be helpful for elucidating the critical question of GRB jets and their structure. Spectro-polarimetry in the radio and IR bands during the early stages of the afterglow, associated with γ-ray polarization measurements, would also place constraints on the inner engine (Sagiv, Waxman, & Loeb 2004). Of course such measurements might already be planned in connection with Swift. They require a rapid decision to observe the early optical afterglow and the early radio flare. The measurement of the polarization of the prompt X- or γ-ray emission will require dedicated instrumentation in space, which will certainly fly in the next 10 years given the high interest in this type of measurement.

9.9.1 Future instruments

A post-Swift mission dedicated to GRBs is already planned for the next decade. This is the Sino-French-Italian mission SVOM which has, among the traditional objectives of a GRB mission (trigger on all GRB types, fast localizations, measurement of the light-curves and spectra, especially \(E_{\text{peak}}\)), the particular desire to focus efforts on the detection and the follow-up observations of nearby and highly redshifted GRBs. This mission includes dedicated instrumentation for the follow-up of detected GRBs: on-board X-ray and visible telescopes, and ground-based robotic telescopes sensitive in the visible and NIR wavelengths.

In the more distant future, one dream would be to have a soft X-ray focusing monitor (capable of detecting very soft bursts from intrinsically faint XRFs and highly redshifted bright GRBs), combined with an NIR telescope in space to identify and follow up the afterglows of nearly all the localized GRBs. Before that dream comes true, the James Webb Space Telescope (JWST; Gardner et al. 2006) will certainly make invaluable and unexpected contributions to the study of GRBs. Other instruments or concepts have been proposed—new ideas and objectives may be found in the symposia held every two years: ‘GRBs in the Swift era’ edited by AIP Conference Proceedings.

A last, general comment may be useful to guide the design of new instruments: it is certainly more interesting to be able to study in detail a few GRBs than to
accumulate many bursts with insufficient data on each of them. We have seen that
even now it is already very difficult to have early, long-duration follow-up observa-
tions of Swift GRBs with a redshift, but they are crucial. This situation is reminis-
cent of the first epoch of the GRB story when we accumulated light-curves and spectra of
the prompt $\gamma$-ray emission, trying to understand their characteristics: periodicities,
spectral hardness, spectral evolution, possible annihilation lines or other lines . . . in
the end without much success. This strategy alone is not the best one. Speaking of
these heroic times, which were really exciting, provides a good opportunity to say that
we wish to dedicate this book to the pioneers of this GRB saga. They are reassembled
in Figure 9.19 (from a special issue of Astrophysics and Space Science 75, 1981) when
they attended a small workshop that G. Vedrenne organized in Toulouse a long time
ago (1979). It is a pleasure to honor them, in particular T. Cline. This is also for the
younger readers an opportunity to discover and/or to see again the pioneers in the
field: a handful of scientists quite excited by these mysterious bursts; we did not
imagine at that time that they were coming from the most distant regions of the
Universe.

It is a great pleasure for us to see that the exciting but anecdotal discovery of $\gamma$-
ray bursts by Klebesadel and his colleagues using unsophisticated $\gamma$-ray detectors has
evolved to become a major field of astrophysics 30 years later. It is also a pleasure to
realize that we have taken part, modestly, in this venture and it is really our good
fortune to have witnessed the extraordinary growth of the field. Without wishing to
be too grandiloquent, it has to be recognized that GRBs are really exciting; these
most powerful explosions in the Universe can reveal the birth of the first massive stars
and their explosions, participating in the birth of our ‘metallic’ Universe. If we exist, it
is thanks to these ‘metals’ (elements beyond H and He) that generations of stars have
produced and liberated through formidable explosions: the SNe and particularly the
SNe Ibc which signal the deaths of very massive stars. Some of these SNe Ibc, the
collapsars, send us strange signals: the GRBs. On the other hand, some GRBs are
probably linked to the coalescence of pairs of compact stars, neutron stars, and black
holes, which populate our Universe for very, very long times when they are isolated,
but which evolve inexorably towards a common death when they are associated in
close binary systems. Here again this catastrophic end manifests itself with a formid-
able explosion releasing gravitational waves which may soon become detectable, and
possibly a GRB. We hope that our excitement and enthusiasm for these GRBs has
been shared by the readers of this GRB story.

9.10 REFERENCES

Abbott, B., et al. (2008a) All-sky search for periodic gravitational waves in LIGO S4 data,
Physical Review D 77, 022001.
Abbott, B. R., et al. (2008b) Implications for the origin of GRB 070201 from LIGO observa-
Abdo, A. A., et al. (2007) Milagro constraints on very high energy emission from short


Index

Achromatic break 102, 105, 122, 124, 286, 317, 364, 369–378, 545
Alfvén turbulence 228, 408–410
Amati relation 121–122
Angular momentum transport 398, 440–441, 444
Angular time 238, 263
Bacodine 72, 74, 88
Band function 120
Baryon load 76, 116, 222, 259, 267, 385, 387, 393, 408, 426, 431
Beaming corrected energy 107, 182, 295, 375, 431, 457–458, 480
Bimodal distribution (durations) 2, 4, 51–52
Calorimetry (of GRBs) 244, 351, 412
Cannonball model 139–140, 250–251, 346
Canonical X-ray afterglow 153, 168, 357
Circularity problem 481, 485
Closure relations 275, 307, 317
Cocoon 299, 336, 394–396
Collisional excitation 501
Collisionless shocks 224, 229, 407, 522, 532–535
Compactness 448–449, 516, 518
Contact discontinuity 261–262, 266–269, 277, 311
Continuous energy injection 312, 338–339, 369
Cooling frequency 139, 144, 189, 272–274, 307, 366, 419
Cosmic microwave background (CMB) 479, 481, 485–486, 520–521
Curvature effect 170, 176, 333–337, 359, 376
Dark ages 185, 477, 486
Dark GRBs 87, 99–102, 140, 153, 416
Deceleration radius 221, 223, 260, 263, 267–268, 315, 337, 357
Dust depletion 489, 491–496, 499
Dust destruction 100, 495
Dust extinction 167, 182, 191, 420, 481, 492–495, 505
Dust extinction law 99–100, 493–496
Efficiency of internal shocks 243–246, 251, 350–351, 376
Electron power-law index 230, 232, 274, 312, 534
Energy-to-mass ratio 221, 259, 393
Equal arrival time surface 288
Equation of state (NS) 362, 426, 432, 436, 440–441, 445–447
Extragalactic background light (EBL) 520

Failed GRB 415, 525
Failed supernova 77, 108, 387, 408, 420
Fall-back 193–194, 362–364, 390, 420, 448, 461
Fermi acceleration 228–229, 242, 477, 513, 522, 526, 534
Fragmentation (core, disk) 359–362, 377, 460, 529

Gravitational lensing 71
Gravitational wave 77, 151, 294, 402, 440, 444, 450, 455, 477, 506, 528–529
GRB 700822 2
GRB 720427 6, 8
GRB 721129 13
GRB 721218 6–7
GRB 730302 7
GRB 730606 7
GRB 730610 7
GRB 730721 7
GRB 730725 7
GRB 731217 7
GRB 731223 7
GRB 771029 4
GRB 780325 12
GRB 780918 3
GRB 781104 13–14
GRB 781119 9–10, 14, 30–32
GRB 790313 3
GRB 790509 2
GRB 790613 2–3, 31–32
GRB 800419 13–14
GRB 811231 14
GRB 820313A 3
GRB 821104 11
GRB 830801B 2–3, 5, 12
GRB 840805 4
GRB 870303 15
GRB 880205 14–15
GRB 890929 15
GRB 910503 76
GRB 910601 54–55, 239
GRB 910711 50
GRB 910814 76, 239
GRB 920311 73
GRB 920325 73
GRB 920501 72
GRB 920517 73
GRB 920622 239
GRB 920723 117–118
GRB 920830 73
GRB 930118 73
GRB 930131 55, 76
GRB 940131 507
GRB 940217 54–55, 233, 507, 512, 517
GRB 940301 73
GRB 941017 507–508
GRB 970111 243
GRB 970402 97
GRB 970417A 234, 507, 516
GRB 970712 416
GRB 970828 95, 99–100, 113, 493
GRB 971110 507
GRB 971214 97–99
GRB 980326 100, 109–111, 289, 416
GRB 980329 100, 302
GRB 980519 103–104, 110, 289, 308, 314, 536
GRB 980613 117, 419
GRB 980703 110, 244, 302, 416
GRB 980923 117
GRB 981121 283
GRB 981223 283
GRB 981226 116
GRB 990510 104–105, 118, 302, 305–307, 310, 314, 373, 536
GRB 990704 116
GRB 990712 110, 116, 302, 305–306
GRB 991208 416, 504
GRB 991216 113, 278, 305, 403, 416, 536
GRB 000131 182, 536
GRB 000210 100
GRB 061210 460
GRB 061217 460
GRB 070129 160
GRB 080319B 283, 539
GRB counterparts 28, 29–33, 40, 71–75, 151
GRB–DLA 491, 497–505
GRB intensity distribution 20–28, 33, 49, 57–61, 64, 298

Hardness ratio 9, 11, 53, 57, 68
Hardness–intensity correlation 55, 57, 121–122, 152, 238
High velocity galactic NS 28, 35
Hugoniots conditions 224
Hypermassive Neutron Star 439–441, 448, 461
Hypernova 98, 109, 113–114, 141, 146, 148, 159, 387, 401, 410–413

Initial energy of the fireball 221
Injection break 367, 369
Innermost stable circular orbit (ISCO) 447–448
Intergalactic magnetic field 520–521
Interplanetary Network (IPN) 18, 20, 31–32, 71–73, 89, 91, 121
Interstellar scintillation 96
Isotropic equivalent energy 104, 106, 112, 124, 141, 159, 184, 189, 286, 293, 298, 351, 369, 385, 413, 438, 457, 480, 483
Isotropy (spatial) 18, 20, 27, 33, 35, 58, 62–66
Jet beaming angle 102–103, 106–107, 122–124, 244, 261, 292, 295–297, 317, 444, 540

Kinetic energy of the ejecta 104, 222, 243, 245, 247, 250, 259, 261, 269, 286, 293–294, 315, 340–344, 461

LAEMW 407, 409
Line-of-death 240–243, 251

Magnetic barrier 361, 377
Magnetic bubble 442–443
Magnetic reconnection 156, 247, 363, 392, 407
Magnetization of the ejecta 284–286, 532–533
Magneto-Rotational Instability 433, 440–441
Mass loss rate 139, 185, 193, 348, 397, 537, 541

Naked GRB 308, 370
Neutrino annihilation 77, 357, 389, 426–452, 462
Neutron rich ejecta 355–356, 511, 517–518, 524
Nickel 56 109, 191, 193, 417–421

Observer timescale 225, 263
Off-axis GRB 152–156, 159, 175, 178, 292, 296, 302, 318, 341, 412, 421, 438, 539
Opacity (pair) 220–222
Opacity (neutrino) 435–436
Opacity (IGM) 507, 520
Optical depth (gamma–gamma) 27, 66, 75, 221–222, 518–520

Pair cascade 407, 409
Pair plasma 66, 75, 406, 445
Particle acceleration 227–228, 242–243, 251, 406, 533
Patchy shell model 139, 175, 310–312, 335, 340–341, 354
Photo-meson interaction 511–512, 515, 521, 526
Photospheric radius 223, 261
Polarization (afterglow) 138, 145, 301–306, 546
Polarization (GRB) 158–159, 532, 546
Population III stars 185–187, 485–489
Proton synchrotron 509–515
Pseudo-redshift 123, 483
Pulsar wind bubble 503
Quark star 357–358, 364, 402, 542
Re-ionization 185, 187, 477, 485–487
Relativistic beaming 76, 102, 220, 246, 484
Robotic telescopes 74, 119, 167, 183, 189, 285, 546
Saturation radius 222
Sedov length 267–268
Shock breakout 180, 393, 413
Sideway expansion 103–104, 287–289, 341
Starquakes 31, 35–36
Steep decay 170, 172, 177, 198, 332–336, 348, 353, 376
Synchrotron frequency 167, 229–231, 240, 274, 280, 337, 367, 513
Thermonuclear model 35–38
Type I GRB 194, 461
Type II GRB 194, 461
Two-component jet 145, 299–300, 318, 341, 343, 354
UMHDW 409
UV pumping 501
\( V/V_{\text{max}} \) 18, 26–28, 35, 38, 57, 60–62, 112
Vela spacecraft 1–2, 6, 30, 33, 38
Viscosity parameter 436, 452
Viscous dissipation 360–361, 429, 429, 432, 436, 441, 445, 452, 460, 462
Visual extinction 493–494
Weibel instability 229, 533–536
XRF 011030 152, 171
XRF 020427 152
XRF 020903 152–153, 160, 177–178, 411, 414, 482
XRF 030429 153
XRF 030528 153
XRF 030723 154, 178, 281, 414, 482
XRF 040701 153, 414, 420
XRF 040812 414
XRF 040912 153
XRF 040916 152, 414
XRF 050406 172, 174
XRF 050408 153
XRF 050416A 152–153, 373–375
XRF 050824 153
Figure 3.1. Images of the source 1 SAX J0501.7+1146, detected with the BeppoSAX Medium Energy Concentrator Spectrometer (2–10 keV) in the error box of GRB970228, during two target of opportunity (TOO) observations. The first observation started only 8 h after the GRBM trigger. The second observation was performed about three days after the GRB. The source faded by a factor of $\sim 20$ in 3 days (Costa et al. 1997e).

Figure 4.1. The HETE-2 satellite during final tests.
Figure 4.2. The network of secondary ground stations receiving the HETE-2 VHF signal (Ricker, private communication).

Figure 4.7. X-ray images of GRB 031203 observed with Epic on XMM-Newton and covering the 0.7–2.5 keV energy range. These observations span a period of time extending from 6 h to 16 h after the burst. The X-ray halo appears as concentric ring-like structures centered on the GRB position (Vaughan et al. 2004).
Figure 4.6. (B) These Hubble Space Telescope images show the fading afterglow and host galaxy of GRB 050709. The images are taken 5.6, 9.8, 18.6, and 34.7 days after the burst, respectively. The bright, point-like afterglow is located to the left, and fades away over the course of the month following the burst. The colors indicate the intensity of red light (814 nm) as seen by the Advanced Camera for Surveys instrument on HST (Fox et al. 2005).

Figure 4.8. Schematic view of the Swift satellite (Gehrels et al. 2004). The size of the coded mask of BAT is 2.7 m².

Figure 4.9. The main elements of the Swift platform (Gehrels et al. 2004).
Figure 5.7. Schematic illustration of the electromagnetic model (EMM) proposed by Lyutikov and Blandford (2003).

Figure 6.1. An artist’s view illustrating the various steps of the basic standard model with the internal and external shocks and the various radiations they emit (from Gehrels, Piro, & Leonard 2007). On the left are indicated the two main scenarios which lead to a central black hole surrounded by a disk. These two types of progenitors are discussed in Chapter 8.
Figure 6.2. Schematic view of the phases involved in the fireball model with internal and external shocks. The left-hand scale gives the logarithm of the distance from the central object producing the GRB (meters). The right-hand scale indicates the various phases: (1) Phase of acceleration: the thermal or magnetic energy is converted into kinetic energy of the baryons within the ejecta. The ejecta emitted in a jet characterized by its beaming angle becomes ultrarelativistic. (2) End of the acceleration phase, the Lorentz factor of the ejecta reaches its maximum value, typically few hundred. (3) The ejecta become transparent. We have already defined the photospheric radius; if internal energy is available it can be radiated as quasi-thermal emission. (4) Internal shocks (CI) appear and progress within the relativistic ejecta. The matter heated by the shocks is the source of the GRB. (5), (6), (7) These phases characterize the various parts of the external shocks, which start when the deceleration due to the circum-burst medium appears: (5) corresponds to the reverse shock (CR) which propagates within the ejecta; (6) is the contact discontinuity (see also Figure 6.3 for the different regions of the external shock); (7) is the forward shock (CA) that is responsible for the afterglow with its emission from radio to gamma-ray wavelengths. (8) Corresponds to a region where the Lorentz factor has greatly decreased and the expansion becomes non-relativistic. Depending on the kinetic energy of the ejecta and the density of the circum-burst medium, the two main phases of internal shocks and external shocks can be well separated or quasi-simultaneous. (Atteia & Mochkovitch 2004).
Figure 8.3. Lorentz factor for the collapsar model $W_2$ of Zhang, Woosley, and MacFadyen (2003) at $t = 10 \text{s}$ (left) and at $t = 35 \text{s}$ (right) after the jets start to propagate in the stellar wind. The initial opening angle is $3^\circ$. At $t = 35 \text{s}$ the opening angle is $5^\circ$. Because the power and Lorentz factor decrease gradually after 10 s the tail of the jet has more lateral expansion.

Figure 6.18. Optical polarization measurements of five GRB afterglows up to 2002. The figure shows the degree of polarization and the corresponding position angle at different times of the afterglow (Covino 2004).
Figure 8.13. Ratio of the energy deposited via neutrino annihilation \((\nu \bar{\nu} \rightarrow e^+ e^-)\) to baryon rest mass energy for the region above the poles of the merged remnant. This is a measure of the maximum attainable Lorentz factor. Due to the centrifugal evacuation of the funnels, the largest Lorentz factors are obtained along the binary rotation axis. At large angles from the rotation axis an increasing degree of entrainment leads to a drastic decrease of the Lorentz factor. Starting from the top three models are considered (from Rosswog, Ramirez-Ruiz, & Davies 2003). (a) Merger of two 1.4 \(M_\odot\) non-rotating neutron stars 18.3 ms after the merger, with neutrino cooling. (b) Merger of two 1.4 \(M_\odot\) corotating neutron stars 10.7 ms after the merger, with neutrino cooling. (c) Merger of two 2.0 \(M_\odot\) non-rotating neutron stars 18.3 ms after the merger, with neutrino cooling.
Figure 8.17. Snapshots of the coalescence of two magnetized NSs followed between 1.83 ms and 11.34 ms. The color scale gives log(\(B\)) (in gauss). The dimensions of the panel are \(\sim 140\) km from left to right. The stars move gradually toward each other and then merge in a ‘plunging phase’ within about one orbital period (\(\sim 2\) ms, the first two snapshots). This new object sheds mass into spiral arms that are subsequently wrapped around the central object (snapshots 3 to 5) to form a hot torus (last snapshot). The magnetic field is amplified in the shear instability between the stars and subsequently advected with the matter to cover the surface of the central merger remnant (from Price & Rosswog 2006).