Lecture Notes in Physics

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Tokyo

Physics and Astronomy ONLINE LIBRARY
http://www.springer.de/phys/

Editorial Policy

The series *Lecture Notes in Physics* (LNP), founded in 1969, reports new developments in physics research and teaching -- quickly, informally but with a high quality. Manuscripts to be considered for publication are topical volumes consisting of a limited number of contributions, carefully edited and closely related to each other. Each contribution should contain at least partly original and previously unpublished material, be written in a clear, pedagogical style and aimed at a broader readership, especially graduate students and nonspecialist researchers wishing to familiarize themselves with the topic concerned. For this reason, traditional proceedings cannot be considered for this series though volumes to appear in this series are often based on material presented at conferences, workshops and schools (in exceptional cases the original papers and/or those not included in the printed book may be added on an accompanying CD ROM, together with the abstracts of posters and other material suitable for publication, e.g. large tables, colour pictures, program codes, etc.).

Acceptance

A project can only be accepted tentatively for publication, by both the editorial board and the publisher, following thorough examination of the material submitted. The book proposal sent to the publisher should consist at least of a preliminary table of contents outlining the structure of the book together with abstracts of all contributions to be included.
Final acceptance is issued by the series editor in charge, in consultation with the publisher, only after receiving the complete manuscript. Final acceptance, possibly requiring minor corrections, usually follows the tentative acceptance unless the final manuscript differs significantly from expectations (project outline). In particular, the series editors are entitled to reject individual contributions if they do not meet the high quality standards of this series. The final manuscript must be camera-ready, and should include both an informative introduction and a sufficiently detailed subject index.

Contractual Aspects

Publication in LNP is free of charge. There is no formal contract, no royalties are paid, and no bulk orders are required, although special discounts are offered in this case. The volume editors receive jointly 30 free copies for their personal use and are entitled, as are the contributing authors, to purchase Springer books at a reduced rate. The publisher secures the copyright for each volume. As a rule, no reprints of individual contributions can be supplied.

Manuscript Submission

The manuscript in its final and approved version must be submitted in camera-ready form. The corresponding electronic source files are also required for the production process, in particular the online version. Technical assistance in compiling the final manuscript can be provided by the publisher's production editor(s), especially with regard to the publisher's own Latex macro package which has been specially designed for this series.

Online Version/ LNP Homepage

LNP homepage (list of available titles, aims and scope, editorial contacts etc.):
http://www.springer.de/phys/books/lnpp/

LNP online (abstracts, full-texts, subscriptions etc.):
http://link.springer.de/series/lnpp/

A. W. Guthmann M. Georganopoulos
A. Marcowith K. Manolakou (Eds.)

Relativistic Flows in Astrophysics

Springer

Editors

Axel W. Guthmann
Markos Georganopoulos
Konstantina Manolakou
Max-Planck-Institut für Kernphysik
69117 Heidelberg, Germany

Alexandre Marcowith
CESR
9, Avenue du Colonel Roche
31028 Toulouse Cedex 4, France

Cover Picture: a compact object (neutron star or black hole) accreting mass from a companion star and sending out a jet. The accretion disc and the corona expected to form can be also seen (see figure 1, page 102, contribution of R. Fender in this volume)

Library of Congress Cataloging-in-Publication Data

Relativistic flows in astrophysics / A.W. Guthmann ... [et al.] (eds.).
p. cm. -- (Lecture notes in physics, ISSN 0075-8450 ; 589)
Includes bibliographycal references and index.
ISBN 3540435182 (acid-free paper)
1. Astrophysical jets. 2. Relativistic astrophysics. I. Guthmann, A. W. (Axel W.), 1973-II. SURF 2000 (2000 : Mykonos, Greece) III. Series

QB466.J46 R45 2002
523.01--dc2l

2002067240

ISSN 0075-8450
ISBN 3-540-43518-2 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Printed in Germany

Typesetting: Camera-ready by the authors/editor
Camera-data conversion by Steingraeber Satztechnik GmbH Heidelberg
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
SPIN: 10867705 54/3141/du - 5 4 3 2 1 0

Prologue

This book contains the contributions of the invited speakers at the conference on *Similarities and Universality in Relativistic Flows* (SURF 2000), which took place from October 1st to 5th, 2000, on the island of Mykonos, Greece. A companion volume, including the contributed talks, has been published by Logos Verlag, Berlin, under the title "Similarities and Universality in Relativistic Flows". The meeting was part of the European Union PhD Euroconference Program and aimed to give young European astrophysicists a platform to present and discuss recent developments in observation and theory of relativistic flows in astrophysical objects.

The topic of this book is the study of relativistic outflows such as those seen in gamma-ray bursts (GRBs), microquasars and active galactic nuclei (AGN). These flows are associated with compact objects ranging from neutron stars and stellar-mass black holes to supermassive black holes ($10^6 - 10^9$ solar masses). The first three contributions in the book deal with the physical processes believed to be important in all families of systems exhibiting relativistic flows. The first one focuses on the relevant radiative processes and Doppler-boosting of the radiation from relativistically moving sources. Then particle acceleration via the first-order Fermi process in relativistic shock fronts is studied, as this mechanism is a good candidate for supplying accelerated electrons responsible for the highly non-thermal spectra being observed. After that, possible jet formation and collimation mechanisms are presented.

The next two contributions present an overview of our understanding of powerful AGN. First, the evolution of classical radio galaxies through studies of their large-scale structure is addressed, followed by a review of our understanding of blazars, which are AGN with jets almost aligned to our line of sight and observed luminosities dominated by a sub-parsec scale region close to the central black hole. Shifting to stellar-size systems, our current understanding of so-called microquasars, binary star systems resembling in many aspects scaled-down versions of AGN, is presented. The next contribution deals with observation and theory of GRBs. The study of the proper bursts and their afterglows suggests the existence of relativistic jets in these objects.

The fluid dynamics of relativistic astrophysical flows are quite complex, and one often has to resort to numerical simulations to describe them. Results of such numerical experiments are presented and compared to recent radio interferomet-

ric observations of jets. This is followed by an introduction to the development of modern numerical techniques for solving relativistic hydrodynamical problems.

The epilogue gives an overview of all the topics discussed at the SURF 2000 conference and offers an outlook over current and future research opportunities.

The meeting was attended by 39 participants from 16 different countries (Australia, Canada, France, Germany, Greece, The Netherlands, Ireland, Israel, Italy, Japan, Norway, Portugal, Russia, Spain, UK, USA). Funding for the conference was provided by the European Union programme on *High Level Scientific Conferences*, under contract HPCFCT-2000-00109. We wish to thank the company Starlab for funding the accommodation and travel expenses of some invited speakers and the administrative staff of the Max-Planck-Institut für Kernphysik for their help in organising the conference. We especially thank the authors of the articles included in this book for their contributions.

Heidelberg,
October 2001

Axel Guthmann
Markos Georganopoulos
Alexandre Marcowith
Konstantina Manolakou

Contents

List of Contributors

Miguel A. Aloy
Max-Planck-Institut für Astrophysik
Karl-Schwarzschild-Str. 1
85748 Garching, Germany
Miguel.A.Aloy@uv.es

Roger Blandford
130–33 California Institute
of Technology
Pasadena CA 91125, USA
rdb@tapir.Caltech.edu

Katherine M. Blundell
Oxford University, Astrophysics Dept.
Keble Road
Oxford
OX1 3RH, UK
kmb@astro.ox.ac.uk

Annalisa Celotti
S.I.S.S.A.
via Beirut 2–4
34014 Trieste, Italy
celotti@sissa.it

Rob Fender
Astronomical Institute
'Anton Pannekoek' and
Center for High Energy Astrophysics
University of Amsterdam
Kruislaan 403
1098 SJ Amsterdam, The Netherlands
rpf@astro.uva.nl

Titus J. Galama
Astronomy 105-24
California Institute of Technology
Pasadena CA 91125, USA
tjg@astro.caltech.edu

Yves A. Gallant
Osservatorio Astrofisico di Arcetri
Largo E. Fermi 5
50125 Firenze, Italy
Y.A.Gallant@astro.uu.nl

José-Luis Gómez
Instituto de Astrofísica de Andalucía
Apartado 3004
18008 Granada, Spain
jlgomez@iaa.es

José M. Martí
Departamento de Astronomía
y Astrofísica
Universidad de Valencia
C/O Doctor Moliner 50
46100 Burjassot, Spain
jose-maria.marti@uv.es

Apostolos Mastichiadis
Department of Astrophysics,
Astronomy and Mechanics
Faculty of Physics
University of Athens
15783 Zografos, Greece
amastich@cc.uoa.gr

Re'em Sari
Theoretical Astrophysics 130-33
California Institute of Technology
Pasadena CA 91125, USA
sari@tapir.caltech.edu

Christophe Sauty
Université Paris 7
Observatoire de Paris, D.A.E.C.
92190 Meudon, France
christophe.sauty@obspm.fr

Edoardo Trussoni
Osservatorio Astron. di Torino
Strada Osservatorio 20
10025 Pino Torinese, Italy
trussoni@to.astro.it

Kanaris Tsinganos
Department of Physics
University of Crete
710 03 Heraklion, Crete, Greece
tsingan@physics.uoc.gr

Radiative Processes in Relativistic Outflows

Apostolos Mastichiadis

Department of Astrophysics, Astronomy and Mechanics, Faculty of Physics,
University of Athens, Zografos 15783, Greece

1 Introduction

During the last decade new, exciting observations shed fresh and sometimes unexpected light on objects generally related with relativistic outflows. Familiar, but still largely unknown[1] objects like Active Galactic Nuclei proved to be strong $\gamma-$ray emitters, the first counterparts of Gamma Ray Bursts were discovered, placing these enigmatic phenomena firmly at cosmological distances, while detection of superluminal motion from X-ray binaries in our own Galaxy showed that these objects act as scaled-down versions of quasars. Depite the fact that many pieces of the puzzle are still missing and obviously the respective differences of these three classes of objects are significant, it is evident that at least at some fundamental level they share some common properties. For instance, there is now consensus that all three classes show some kind of relativistic outflow and relativistic beaming has been invoked to relax the constraints imposed especially on luminosities and high energy photon opacities. Furthermore, non-thermal radiation processes such as synchrotron and inverse Compton scattering, seem to be the radiation mechanisms that are mostly responsible for the sources' spectral formation.

In the present paper, we will make a textbook-like review of the phenomena and of the radiative processes that operate in these objects. We will begin by reviewing the phenomenon of superluminal motion. Next we will present the way Doppler boosting operates and give arguments that this in combination with $\gamma-$ray opacities favor relativistic outflows. Then we will briefly present synchrotron radiation and inverse Compton scattering. We will use these processes to demonstrate how one can use simple forms of the continuity equation to model variable emission that is quite common in all classes of sources. Finally, we will finish with a brief note on the role relativistic hadrons might play in these objects.

It is evident that so many different topics cannot easily fit in an article. Thus I will not attempt to make a comprehensive review of the subjects, but I will simply introduce the concepts giving emphasis on these aspects that play the role of theoretical tools to the astrophycisists. I will take a similar approach with the references: As my aim is not an updated review of classes of objects or of proposed models, but rather of mechanisms, I will mainly refer to some older key papers. This unfortunately will leave out some more recent papers which

[1] apologies for the $o\xi\acute{\upsilon}\mu\omega\rho o\nu$

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 1–23, 2002.

have applied and partly extended some of the ideas presented here. There will be only a few exceptions to this rule, mainly in the last sections.

It is hoped that this generic, model-independent approach will help the reader to refresh some of the issues and can be used as general background for the more detailed articles to be found later in this volume.

2 Superluminal Motion

Superluminal motion is a direct manifestation of Einstein's special theory of relativity. Its relevance to astrophysical objects has been predicted [25] about five years before observations [4] demonstrated its existence. The effect is based on the fact that if some blob of photon emitting plasma, which is initially at a distance D from us, is moving away from a central 'core' with velocity $\upsilon = \beta c$ making an angle θ to our line of sight, then the photons which are emitted at a later time t_e will not cover the distance D to us but rather the distance $D - \upsilon t_e \cos\theta$. This has the immediate consequence that the time interval Δt_{obs}, which we measure between photon emission instances, is shortened by the factor $1 - \beta\cos\theta$ over the corresponding time interval as measured in the blob rest frame. As a result the apparent velocity of the blob on the plane of the sky is measured as

$$\upsilon_{app} = \frac{\upsilon t_e \sin\theta}{\Delta t_{obs}} = c\frac{\beta\sin\theta}{1-\beta\cos\theta}. \tag{1}$$

This quantity can be greater than 1 when the denominator becomes small, which means that both β and $\cos\theta$ should approach unity. In this case the above relation can be written

$$\beta_{app} = \frac{\upsilon_{app}}{c} \simeq \frac{2\theta}{\Gamma^{-2}+\theta^2} \tag{2}$$

where Γ is the Lorentz factor[2] of the blob and we have used $\beta \simeq 1 - 1/2\Gamma^2$ and $\cos\theta \simeq 1 - \theta^2/2$. In the case where $\Gamma^{-1} < \theta \ll 1$, we get $\beta_{app} \simeq 2\theta^{-1} \gg 1$.

Another way to see this is to solve Eqn (1) for β, i.e.

$$\beta = \frac{\beta_{app}}{\sin\theta + \beta_{app}\cos\theta} \tag{3}$$

and since $\beta < 1$, we arrive by solving the resulting inequality at the relation

$$\frac{\beta_{app}^2 - 1}{\beta_{app}^2 + 1} < \cos\theta < 1. \tag{4}$$

When $\beta_{app} > 1$, the lower limit of the above inequality gets close to 1, and therefore the angle θ can take only small values.

The maximum value which β_{app} can achieve is obtained for $\cos\theta = \beta$ and in that case

$$\beta_{app}^{max} = \frac{\beta}{\sqrt{1-\beta^2}} \tag{5}$$

[2] $\Gamma = 1/\sqrt{(1 - v^2/c^2)}$ – *the eds.*

which means that, under optimal conditions, superluminal motion can be produced with velocities as low as $.7c$.

3 Doppler Boosting

Another consequence of relativistic kinematics is the phenomenon of Doppler boosting. We assume, as before, that the source of radiating particles is moving with velocity $\upsilon = \beta c$ and is making an angle θ to our line of sight. The question one would like to ask is how the various observationally inferred source parameters are related to the ones intrinsic to the source. [19] have described the phenomenon in the case of different relativistic jet models. Here we will adopt a different method as discussed in [26].

We start with the definition of the flux $S_\nu = \int I_\nu d\Omega$ and we assume that the source has a thickness s, is optically thin and at a distance D from us. Then by using the relations $d\Omega = dA/D^2$, $I_\nu = j_\nu s$, $dV = dA.s$ we rewrite the flux as

$$S_\nu = \int j_\nu dV/D^2. \tag{6}$$

Therefore we need to know how the emission coefficient in the observer's frame j_ν is related to $j'_{\nu'}$ ($'$ denotes the rest frame of the flow). For this we use the definition of the emission coefficient

$$j_\nu = n\frac{dW}{dtd\Omega d\nu} \tag{7}$$

where n is the density of emitters and we will transform each of these quantities. To do this we define the *Doppler factor*

$$\delta = \Gamma^{-1}(1 - \beta\cos\theta)^{-1} \tag{8}$$

where Γ is the bulk Lorentz factor of the flow and we have for each of the quantities in (7):

- Frequency $d\nu' = \delta^{-1}d\nu$ (Doppler formula)
- Energy $dW' = \delta^{-1}dW$
- Time $dt' = \Gamma^{-1}dt$
- Density of emitters $n' = \Gamma^{-1}n$
- Solid angle $d\Omega' = \delta^2 d\Omega$
  (because $\frac{d\Omega\prime}{d\Omega} = \frac{d\cos\theta'}{d\cos\theta}\frac{d\phi'}{d\phi}$ and $\cos\theta' = \frac{\cos\theta-\beta}{1-\beta\cos\theta}$)

Putting all the above relations in (7) we get

$$j'_{\nu'} = \delta^{-2}j_\nu. \tag{9}$$

With the help of the above relation, Eqn (6) becomes

$$S_\nu(\nu) = \delta^3 \int j'_{\nu'}(\nu')dV'/D^2. \tag{10}$$

In the case where the emission coefficient in the rest frame of the flow is a power-law, i.e. $j'_{\nu'} \propto (\nu')^{-\alpha}$ we get

$$S_\nu(\nu) = \delta^{3+\alpha} \int j'_{\nu'}(\nu) dV'/D^2 \tag{11}$$

and if the emitting source is spherical an integration over volume and over frequency gives that the observed luminosity is related to the intrinsic luminosity by the relation

$$L = \delta^4 L' \tag{12}$$

which makes it a sharp function of the Doppler factor.

The Doppler factor itself is a very sensitive function of the angle θ. Thus depending on the possible values of this parameter one can show from (8):

- For $\theta < \Gamma^{-1} \rightarrow \delta \simeq \Gamma$ (13)
- For $1 < \theta \rightarrow \delta \simeq \Gamma^{-1}$ (14)

A combination of the relations above with Eqn (12) leads us to the conclusion that relativistically moving sources are boosted when moving towards the observer who deduces a luminosity that can be many orders of magnitude higher that the intrinsic luminosity of the source. Similarly, when the source is moving at large angles to the observer it appears much weaker than it really is.

Extra evidence for Doppler boosting (and therefore relativistic outflows) is provided by the $\gamma-$ray observations of blazars and Gamma Ray Bursts (GRBs). In the next section we present the relevant evidence, after introducing briefly the process of photon-photon pair production.

4 Photon-Photon Pair Production

This is the process where two photons collide and they produce an electron-positron pair. It has a threshold as the energy of the photos in the center-of-momentum frame should be at least equal to the rest mass of the produced particles, thus the condition

$$\epsilon_2 \geq \frac{2(m_e c^2)^2}{\epsilon_1 (1 - \cos\theta)} \tag{15}$$

should be satisfied. Here ϵ_1 and ϵ_2 are the energies of the two photons while θ is the angle of collision. Thus, for example, X$-$ray photons of energy $\epsilon_1 \simeq$ 1 keV can absorb $\gamma-$ray photons of energy $\epsilon_2 \geq 300$ MeV, 10 keV photons can absorb 30 MeV photons and so on. However this is not only a question of energy threshold, but of opacity as well. Since both blazars and GRBs show high luminosity emission both in the X$-$ and in the $\gamma-$ray regime, the question of photon-photon opacity becomes central in understanding the properties of these sources.

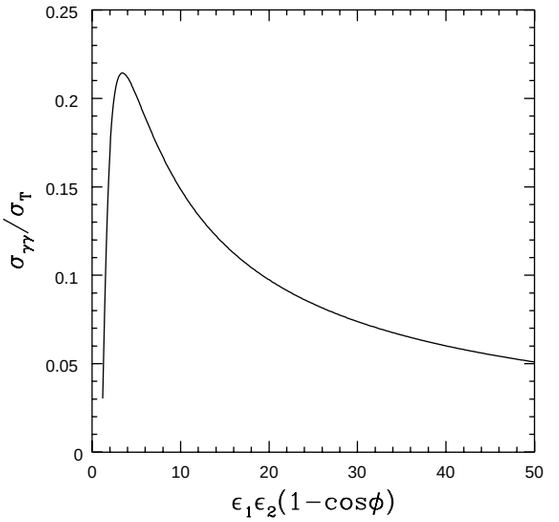

Fig. 1. The cross section of the process $\gamma\gamma \to e^+e^-$ (in units of the Thomson cross section σ_T) as a function of $\epsilon' = \epsilon_1\epsilon_2(1-\cos\phi)$ – the energies are in units of m_ec^2.

Indeed the optical depth for a photon of energy ϵ_2 to cross a region of 'radius' R filled with photons of energy ϵ_1 and differential number density $n(\epsilon_1)$ is given by (see [9])

$$\tau_{\gamma\gamma}(\epsilon_2) = \frac{R}{4\pi}\int d\epsilon_1 n(\epsilon_1)\int d\Omega(1-\cos\theta)\sigma_{\gamma\gamma}(\epsilon') \tag{16}$$

where $\epsilon' = (\epsilon_1\epsilon_2/m_e^2c^4)(1-\cos\theta)$ and $\sigma_{\gamma\gamma}$ is the photon-photon pair production. This is given in Fig. 1 in units of the Thomson cross section $\sigma_T = 8\pi r_0^2/3 = 6.65\times10^{-25}$ cm^2. In the case of an isotropically emitting, optically thin source the photon number density can be deduced from the luminosity according to the relation

$$L = 4\pi R^2 c\int d\epsilon\ \epsilon n(\epsilon). \tag{17}$$

Furthermore we can construct the so-called 'compactness' of the source [10]

$$l = \frac{L\sigma_T}{4\pi R m_e c^3} \tag{18}$$

where σ_T is the Thomson cross-section. As this parameter is related to observable quantities it can be immediately estimated. Thus using $R \simeq c.\Delta t$, where Δt is a typical variation timescale of the source, we have

- AGN: $L = 10^{48}$ erg/sec, $\Delta t = 1$ day $\Rightarrow l \simeq 10^3$
- GRB: $L = 10^{52}$ erg/sec, $\Delta t = 1$ sec $\Rightarrow l \simeq 10^{12}$

Assuming that the photon number density is characterized by a power law of the form $n(\epsilon) = n_0\epsilon^{-s}$ for $\epsilon_{min} \le \epsilon \le \epsilon_{max}$ and further assuming a $\delta-$function

approximation to the photon-photon pair production (choice that is partly justified from the functional form of $\sigma_{\gamma\gamma}$, c.f. Fig 1), we obtain from Eqn (16) (after replacing ϵ_2 with E_γ)

$$\tau_{\gamma\gamma}(E_\gamma) \simeq \frac{1}{4}\eta(\epsilon_{min}, \epsilon_{max}, s)\left(\frac{E_\gamma}{m_e c^2}\right)^{s-1} l \tag{19}$$

where

$$\eta = \begin{cases} (2-s)/(\epsilon_{max}^{2-s} - \epsilon_{min}^{2-s}) & s \neq 2 \quad (20a) \\ [\ln(\epsilon_{max}/\epsilon_{min})]^{-1} & s = 2 \quad (20b) \end{cases}$$

Using the above derived values for l and taking a 'canonical' spectral slope $s = 2$ (which implies $\eta \simeq .1$, a value which is rather insensitive to the limits ϵ_{min}, ϵ_{max}) we get from (19)

- AGN: E_γ= 10 GeV $\rightarrow \tau_{\gamma\gamma} \approx 10^4$
- GRB: E_γ= 10 MeV $\rightarrow \tau_{\gamma\gamma} \approx 10^{12}$

which clearly states that no $\gamma-$rays should have been observed were the photon emission isotropic.

However, if the emission is not isotropic but beamed instead, then the above values are largely relaxed. Indeed in the case where the source is moving relativistically with Doppler factor δ, then the following facts influence directly our calculations of the optical depth:

- The source appears more luminous than what it really is: $L_{obs} = \delta^4 L_{int}$
- The source appears more compact: $\Delta t_{obs} = \delta^{-1}\Delta t_{int}$
- The observed photons appear more energetic: $E_{\gamma,obs} = \delta E_{\gamma,int}$

With the above in mind we can write (denoting now by 'obs' and 'int' the quantities in the observer's and the source's frame respectively):

$$l_{obs} \propto \frac{L_{obs}}{\Delta t_{obs}} = \frac{\delta^4 L_{int}}{\delta^{-1}\Delta t_{int}} \propto \delta^5 l_{int} \tag{21}$$

and consequently the optical depth becomes (for $s = 2$)

$$\tau_{\gamma\gamma,obs} \propto l_{obs} E_{\gamma,obs} = (\delta^5 l_{int}).(\delta E_{\gamma,int}) \propto \delta^6 \tau_{\gamma\gamma,int} \tag{22}$$

For $\gamma-$rays to escape the source we need

$$\tau_{\gamma\gamma,int} = \frac{\tau_{\gamma\gamma,obs}}{\delta^6} < 1. \tag{23}$$

Thus these simplified calulations imply at once that for AGN $\delta > 5$ while for GRB $\delta > 100$. Of course one can do more sophisticated calculations, however the above results will not change much. Therefore the above arguments give indirect but strong evidence that in both blazars and GRBs the emission comes from relativistically moving matter.

5 Synchrotron Radiation

When relativistic electrons spiral along magnetic field lines they accelerate and thus radiate. This process is called synchroton radiation and has been treated in many texts - see, for example [7], [3], [26]. The electrons gyrate about the magnetic field direction with the so called relativistic gyrofrequency:

$$\nu_g = \frac{eB}{\gamma m_e c}, \tag{24}$$

where γ is the Lorentz factor of the electrons and B the magnetic field strength.

The emission is broad-band and consists of overtones of ν_g with a peak close to the so-called critical frequency

$$\nu_c = \frac{3\gamma^2 eB \sin a}{4\pi m_e c} \tag{25}$$

where a is the pitch angle– i.e. the constant angle between the electron velocity and the magnetic field direction . This happens because the emission lies in a cone of angle $\sim \gamma^{-1}$ and, at the same time, the observed pulse width is compressed over the rest frame one by the factor $1 - \frac{v}{c} \simeq \gamma^{-2}$. Combining the above results and using the fundamental relation that the maximum frequency of the spectrum is related to the pulse width by $\omega \sim 1/\Delta t$ one can obtain relation (25).

Much more detailed calculations (see [13]) show that the spectrum radiated from an electron is given by

$$j_s(\nu) = \frac{\sqrt{3} e^3 B sin a}{m_e c^2} F\left(\frac{\nu}{\nu_c}\right) \tag{26}$$

where the function $F(x)$ is defined by $F(x) = x \int_x^\infty K_{5/3}(y) dy$ with $K_{5/3}$ the modified Bessel function of order 5/3. This function has the asymptotic forms

$$\text{For} \quad x \ll 1: \qquad F(x) \simeq \frac{4\pi}{3^{1/2} 2^{1/3} \Gamma(1/3)} x^{1/3} \tag{27}$$

$$\text{For} \quad x \gg 1: \qquad F(x) \simeq \left(\frac{\pi}{2}\right)^{1/2} e^{-x} x^{1/2} \tag{28}$$

while it peaks for $x \simeq 0.3$.

The power radiated by the synchrotron process can be derived in a straightforward fashion using Larmor's formula and is given (after averaging over pitch angles) by

$$P_{syn} = -\left(\frac{dE_e}{dt}\right)_{syn} = \frac{4}{3} \sigma_T c \beta^2 \gamma^2 U_B \simeq \frac{4}{3} \sigma_T c \gamma^2 U_B \tag{29}$$

where $U_B = B^2/8\pi$ is the magnetic energy density. From the above relation one can deduce that the energy loss timescale for the radiating electrons is

$$\tau_{syn} = \frac{E_e}{dE_e/dt} \propto \gamma^{-1} B^{-2}, \tag{30}$$

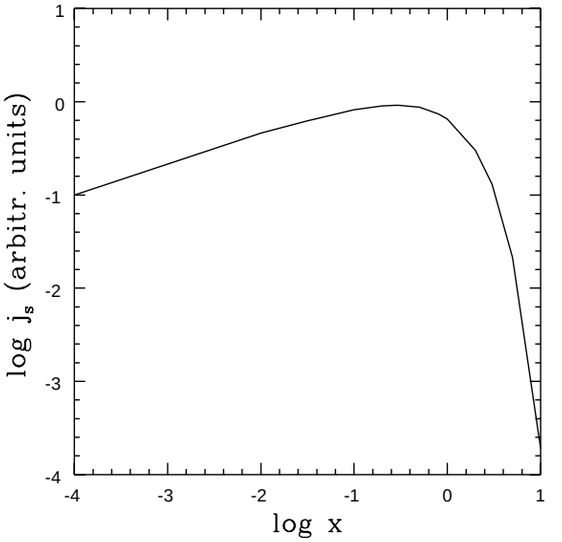

Fig. 2. Single electron emissivity as a function of the parameter $x = \nu/\nu_c$ where ν_c is the critical synchrotron frequency.

which states that higher energy electrons lose energy faster. This will become particularly useful in a next section when we study the time-dependence of synchrotron radiation.

In the astrophysically interesting case where the electrons have a power law distribution , then one can convolve the single electron emissivity over the electron distribution to get the resulting photon spectrum. Fig. 3 shows schematically how this works. Electrons with higher energies will create higher energy[3] synchrotron specta which due to Eqn (25) will be spread by a factor of two in logarithm over the corresponding electrons. Furthermore these spectra have to be convolved with the electron power-law, so finally the photon spectrum will be a power law formed by the common envelope of the single electron emissivities. Despite the fact that exact calculations [7] show that the radiated photons have a spectral index of $(p-1)/2$ where p is the electron index, it is better for the purposes of the present review to show this result in a more approximate way.

Therefore one can follow [11] and assume that the electron is emitting all of its radiation at some characteristic energy ϵ_0 and approximate the single electron emissivity with a $\delta-$function, i.e.

$$j_s = A\epsilon_0\delta(\epsilon - \epsilon_0) \tag{31}$$

where the normalization constant A is taken in such a way as to ensure that the total emitted power is given by Eqn (29). Usually ϵ_0 is taken to be equal to the critical synchrotron energy ϵ_c–see relation (25). However, sometimes it is more

[3] From now on photon energies rather than photon frequencies will be used

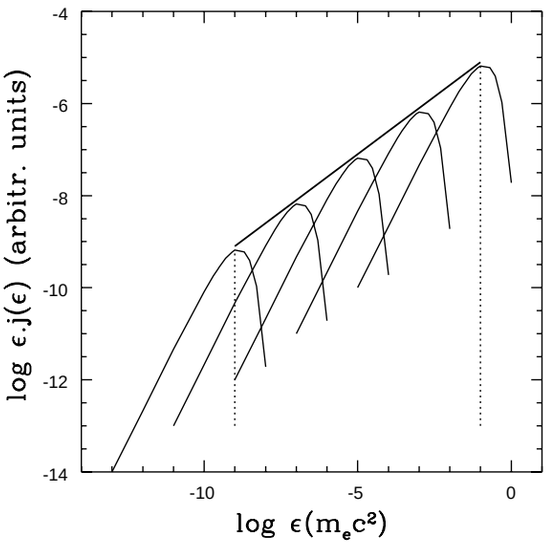

Fig. 3. Schematic representation of the formation of a power-law synchrotron photon spectrum from a power-law electron distibution. Here the emissivity $j(\epsilon)$ is multiplied by the photon energy ϵ and this quantity, when plotted against ϵ, shows where most of the radiated power is going to.

convenient to assume that

$$\frac{\epsilon_0}{m_e c^2} = b\gamma^2 \tag{32}$$

where $b = B/B_c$ and $B_c = m_e^2 c^3/(e\hbar)$ is the critical magnetic field, an approach that we will adopt here – for more on this see [21].

Assume next that the electron distribution function is of the form $N_e(\gamma) = K_e \gamma^{-p}$ and that this holds for $\gamma_{min} \leq \gamma \leq \gamma_{max}$. Then the emissivity due to these electrons will be

$$j(\epsilon) = \int_{\gamma_{min}}^{\gamma_{max}} d\gamma j_s(\epsilon, \gamma) N_e(\gamma) = \int_{\gamma_{min}}^{\gamma_{max}} d\gamma A \epsilon_0 \delta(\epsilon - \epsilon_0) K_e \gamma^{-p} \tag{33}$$

and performing the integral we find with the help of relation (32)

$$j(\epsilon) \propto \epsilon^{-\frac{(p-1)}{2}} \tag{34}$$

which holds for $\epsilon_{min} \equiv \epsilon_c(\gamma_{min}) < \epsilon < \epsilon_{max} \equiv \epsilon_c(\gamma_{max})$ in qualitative agreement with the exact calculations.

The final aspect of synchrotron radiation to be mentioned briefly here is that of synchrotron self absorption. This occurs when the electrons become opaque to their own radiation and it has important consequences for the photon spectral shape. Defining the frequency-dependent brightness temperature

$$T_B = \frac{c^2 I_\nu}{2 k_B \nu^2} \tag{35}$$

where I_ν is the specific intensity and k_B is Boltzmann's constant and the kinetic temperature of the electrons from the relation

$$3k_B T_e = \gamma m_e c^2, \tag{36}$$

we expect that the source will become self-absorbed when $T_B \leq T_e$, i.e. when the electron temperature is to exceed the temperature of the black body of intensity I_ν, which obviously is not permitted. The spectrum below the self-absorption frequency is given by $j(\nu) \propto \nu^{5/2}$, independent of the underlying electron spectrum.

Other, equally important aspects of synchrotron radiation, such as its polarization will not be addressed here. However the reader is referred either to the references given in the beginning of the section or to [2] which is somehow similar in scope with the present review.

6 Inverse Compton Scattering

This process most probably plays an important role in the formation of the high energy spectra of many astrophysical sources. It involves the scattering of ambient, low energy photons to high energies by relativistic electrons and therefore it serves both as a source of hard photons and as an electron energy loss mechanism (see [14]).

The total cross section (known as the 'Klein-Nishina' cross section) of the process depends on the quantity

$$\epsilon' = \gamma\epsilon(1 - \beta\cos\theta) \tag{37}$$

that is the incoming photon energy in the rest frame of the electron (ERF). Here γ is the Lorentz factor of the electron, ϵ is the incoming photon energy in the lab frame and θ is the angle of collision. Then we can write

$$\sigma_{KN} = \frac{3}{4}\sigma_T \left[\frac{1+x}{x^3}\left(\frac{2x(1+x)}{1+2x} - \log(1+2x)\right) + \frac{1}{2x}\log(1+2x) - \frac{1+3x}{(1+2x)^2}\right] \tag{38}$$

where $x = \epsilon'/m_e c^2$. Fig. 4 represents a plot of the total cross section as a function of the incoming photon energy in the ERF. There are two particularly useful asymptotic expressions to σ_{KN}:

- For $x \ll 1 \rightarrow \sigma_{ics} \simeq \sigma_T$
  Thomson regime – classical limit – elastic scattering (ERF)
- For $x \gg 1 \rightarrow \sigma_{ics} \simeq \frac{3}{8}\sigma_T x^{-1}(\ln 2x + \frac{1}{2})$
  Klein-Nishina regime – quantum limit – electron recoil is important (ERF)

Thus when the energy of the collision becomes greater than $m_e c^2$, the cross section decreases and this has important consequences for the application of the process to astrophysical environments.

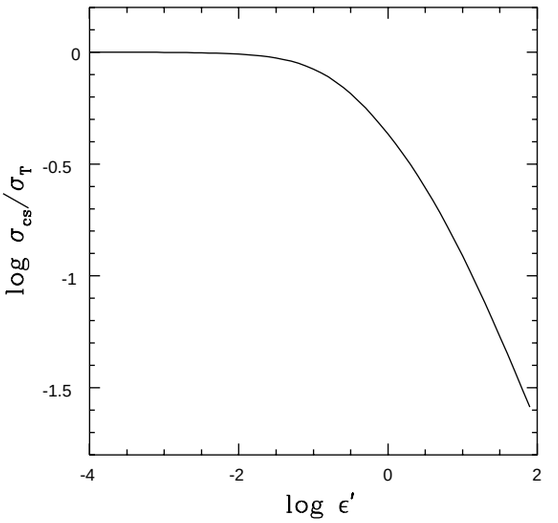

Fig. 4. The cross section for Compton scattering (in units of the Thomson cross section σ_T) as a function of the incoming photon energy as measured in the electron rest frame $\epsilon' = \gamma\epsilon(1 - \beta\cos\theta) - \epsilon'$ is in units of $m_e c^2$.

The scattered photon distribution function in the lab frame in the case of an isotropic distribution of target photons with differential number density $n(\epsilon)$ is given by (see [3])

$$\frac{dN}{dtdE_1} = \frac{3}{4}\frac{\sigma_T m_e c^3}{\gamma}\frac{n(\epsilon)d\epsilon}{\epsilon}F_{ICS}(\epsilon,\gamma) \tag{39}$$

where $E_1 = \epsilon_1/(\gamma m_e c^2)$ and the function F_{ICS} is given by

$$F_{ICS}(\epsilon,\gamma) = 2q\ln q + (1+2q)(1-q) + \frac{1}{2}\frac{(G_e q)^2}{1+G_e q}(1-q) \tag{40}$$

with $G_e = 4\epsilon\gamma/m_e c^2$ and $q = E_1/G_e(1-E_1)$. The parameter G_e defines the scattering regime. For $G_e \ll 1$ this is in the Thomson regime, while for $G_e \gg 1$ it is in the Klein-Nishina regime. Fig. 5 depicts the quantity F_{ICS} as a function of the normalized scattered energy $E_1^* = E_1(1+G_e)/G_e$ for various values of the parameter G_e. As G_e increases, F_{ICS} tends to peak at higher values which means that the scattered photons tend to gain increasingly higher fraction of the electron energy. We will return later to this point, after we examined the opposite case, i.e. collisions in the classical (or Thomson) limit.

6.1 Thomson Scattering

First we derive, using simple arguments, the maximum energy the scattered photon can achieve. For given incoming photon and electron energies, the photon

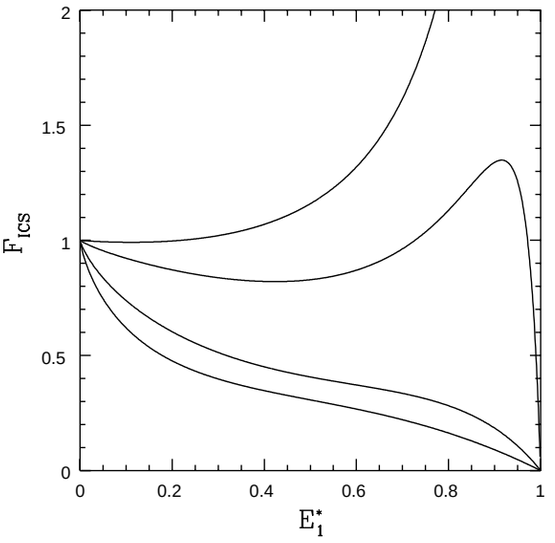

Fig. 5. Plot of the scattered photon distribution function F as a function of the normalized scattered photon energy E_1^* for various values of the parameter G_e. Here G_e takes the values 0, 1, 10 and 100 (from bottom to top).

energy ϵ' as measured in the electron rest frame becomes maximum for head-on collisions ($\theta = \pi$). In that case $\epsilon'_{max} \simeq 2\gamma\epsilon$. As it is known from Compton scattering the energy of the scattered photon (in the electron rest frame) is given by

$$\epsilon'_1 = \frac{\epsilon'}{1 + (\epsilon'/m_e c^2)(1 - \cos\theta'_1)} \tag{41}$$

where θ'_1 is the scattering angle (the angle between the photon direction before and after the scattering). Since we have assumed that the scattering occurs in the Thomson regime we have $\epsilon' \ll m_e c^2$ and relation (41) gives $\epsilon'_1 \simeq \epsilon'$, i.e. the scattering is elastic (in the ERF). If we transform back to the lab frame and use the above relation we obtain

$$\epsilon_1 = \gamma\epsilon'_1[1 + \beta\cos(\pi - \theta'_1)] \simeq \gamma\epsilon'(1 - \cos\theta'_1) \tag{42}$$

which becomes maximum (for given ϵ') when $\theta'_1 = \pi$, so $\epsilon_1 \simeq 2\gamma\epsilon'$. Thus the maximum value the scattered photon energy ϵ_1 can take is when ϵ' itself becomes maximum, so finally we get

$$\epsilon_{1,max} \simeq 4\gamma^2\epsilon \tag{43}$$

which is the result we have been seeking.

Using the above result and the condition $G_e \ll 1$ one can slightly simplify Eqn (40) to

$$\frac{dN}{dt d\epsilon_1} = \frac{3}{16}\frac{\sigma_T c}{\gamma^4}\frac{n(\epsilon)d\epsilon}{\epsilon^2} f(\epsilon_1; \epsilon, \gamma) \tag{44}$$

where

$$f(\epsilon_1;\epsilon,\gamma) = 2\epsilon_1 \log\frac{\epsilon_1}{4\gamma^2\epsilon} + \epsilon_1 4\gamma^2\epsilon - \frac{\epsilon_1^2}{2\gamma^2\epsilon}. \tag{45}$$

An interesting corollary of the above relation is that the average energy of the scattered photon is

$$< \epsilon_1 >= \frac{4}{3}\gamma^2\epsilon \tag{46}$$

Indeed, normalizing ϵ_1 to $\hat{\epsilon}_1 = \epsilon_1/4\gamma^2\epsilon$ and calculating the moments of the function f one gets

$$\int f(\hat{\epsilon}_1)d\hat{\epsilon}_1 = \frac{1}{3} \tag{47}$$

and

$$\int f(\hat{\epsilon}_1)\hat{\epsilon}_1 d\hat{\epsilon}_1 = \frac{1}{9} \tag{48}$$

from which relation (46) immediately is derived.

The electron energy losses in the Thomson regime can also be derived in a straightforward manner. Thus from the general definition

$$-\left(\frac{dE_e}{dt}\right)_{ics} = \int d\epsilon_1(\epsilon_1 - \epsilon)\frac{dN}{dtd\epsilon_1} \tag{49}$$

one can derive, by assuming $\epsilon_1 >> \epsilon$, using relations (44), (45) and (48) and performing the integration over ϵ

$$P_{ics} = -\left(\frac{dE_e}{dt}\right)_T = \frac{4}{3}\sigma_T c\beta^2\gamma^2 U_{ph} \tag{50}$$

where $U_{ph} = \int d\epsilon\epsilon n(\epsilon)$ is the target photon energy density.

Relations (29) and (50) show that there is a close analogy between synchrotron radiation and inverse Compton scattering in the Thomson regime. Indeed both mechanisms cause electron losses that depend on the square of electron energy and therefore are more important loss mechanisms for high energy electrons than, say, bremsstrahlung which depends linearly on the electron energy. Furthermore, both mechanisms depend on the energy densities of either the magnetic field or the target photon field. Therefore in the case where both mechanisms operate on the same relativistic electrons we find that their relative importance is simply the ratio of the two energy densities, i.e.

$$\frac{P_{syn}}{P_{ics}} = \frac{U_B}{U_{ph}}. \tag{51}$$

The analogy between the two mechanisms does not stop there. Comparing relations (32) and (46) we deduce that both mechanisms produce photons that depend on the square of the electron energy. Having in mind that in astrophysical sources usually $b = B/B_{crit} \ll \epsilon/m_e c^2$ one can understand the reason inverse Compton scattering is assumed to produce much more energetic photons than synchrotron radiation.

Finally we note that in the case where the electrons are characterized by a power-law distribution function of the form $N(\gamma) \propto \gamma^{-p}$ the scattered photon spectrum is also a power law of spectral index $a = (p-1)/2$. The arguments leading to this conclusion are completely analogous to the ones given in the previous section for the case of power-law spectra due to synchrotron radiation.

6.2 Klein-Nishina Limit

As we have already mentioned in this limit the relation $\epsilon' \gg m_e c^2$ holds and the scattering in the electron rest frame is no longer elastic but the electron gets a large recoil. Following essentially the same steps as we did for the case of the Thomson limit one can show that the maximum energy photons can achieve, as measured in the lab frame, is of order of $\gamma m_e c^2$, therefore electrons tend to lose most of their energy in one collision, i.e. losses are catastrophic and this, by itself, is a major difference between scattering in the two regimes. However, when one is interested in the electron energy losses, this effect is compensated by the fact that the collision rate is greatly suppressed due to the cross section decrease. Thus starting from the relation

$$\begin{aligned} -\left(\frac{dE}{dt}\right)_{KN} &= (\text{Scattering rate}) \times (\text{Average energy transfer/scattering}) \\ &= \int [d\epsilon n(\epsilon)\sigma_{KN}(\gamma\epsilon)c]\,(\gamma m_e c^2) \end{aligned} \tag{52}$$

we can write to first approximation

$$P_{KN} = -\left(\frac{dE}{dt}\right)_{KN} \simeq \frac{3}{8}\sigma_T m_e^2 c^5 \int d\epsilon \frac{n(\epsilon)}{\epsilon}\left[\log\frac{2\gamma\epsilon}{m_e c^2} + \frac{1}{2}\right] \tag{53}$$

More exact expressions can only be derived in a numerical fashion and involve calculation of the integral in relation (49). The final result is not very different from the qualitative one obtained above and indicates that inverse Compton scattering becomes less efficient at higher energies. Fig. 6 shows the impact of the Klein-Nishina effect on the electron energy losses. Here an isotropic monochromatic target photon field has been assumed and the losses are calculated as a function of the electron energy. As long as the scattering is in the Thomson regime the losses have the expected γ^2 dependence. However, as the electron energy increases and scatterings start predominantly occuring in the Klein-Nishina regime, there is a marked steepening of the losses which, at the extreme Klein-Nishina regime, have only a logarithmic dependence on the electron energy.

An immediate result of the above is that the scattered photon spectrum produced from a power-law electron distribution will be steeper than if the scatterings were taking place in the Thomson regime. Indeed in this case one can show that the photon spectrum will be $\propto \epsilon_1^{-p}$, where p is the electron index.

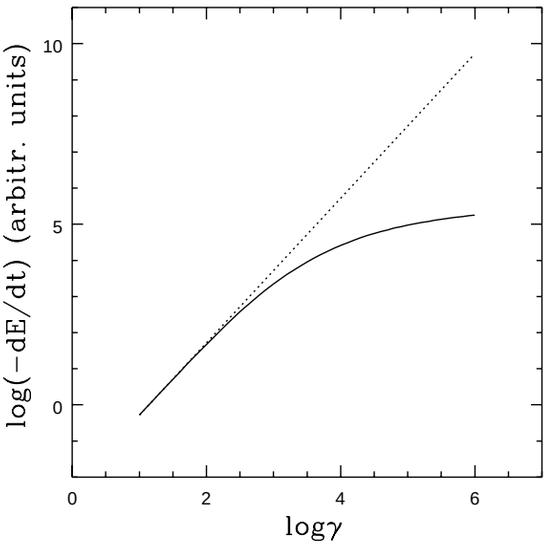

Fig. 6. Inverse Compton losses for monoenergetic electrons as a function of the Lorentz electron factor for an isotropic monoenergetic soft photon field of characteristic energy of $10^{-3} m_e c^2$.

7 Synchro Self-Compton Radiation

When inverse Compton scattering is applied to specific astrophysical objects the question of the origin of the target photons becomes a central one. These could come, for instance, from sources independent of the radiating electrons, such as accretion disks, emission line clouds, etc – these are the so-called external photon models. However, another possibility is that the target photons could be produced by the electrons themselves. A mechanism that can produce the required soft photons is synchrotron emission and in this case the process is called synchro self-Compton (SSC) radiation. The process was originally proposed to explain X-ray emission from Active Galactic Nuclei [15] and was revisited to explain the $\gamma-$ray emission from blazars.

There are a few features which distinguish SSC emission from external inverse Compton scattering. Perhaps the best known one is that as the synchrotron emissivity depends linearly on the electron distribution function, the SSC emissivity will depend, on the one hand, on the same electron distribution function and, on the other, on the synchrotron emissivity. Therefore in steady state and assuming that all scatterings occur in the Thomson regime, any change in the normalization of the electrons will cause linear changes in the synchrotron component and quadratic in the SSC component as is depicted in Fig. 7.

Also the limits of the photon distribution can be found in a straightforward manner. Thus since, according to Section 5, the synchrotron component is emitted between $\epsilon_{s,min} \simeq b\gamma_{min}^2$ and $\epsilon_{s,max} \simeq b\gamma_{max}^2$, the SSC component will extend

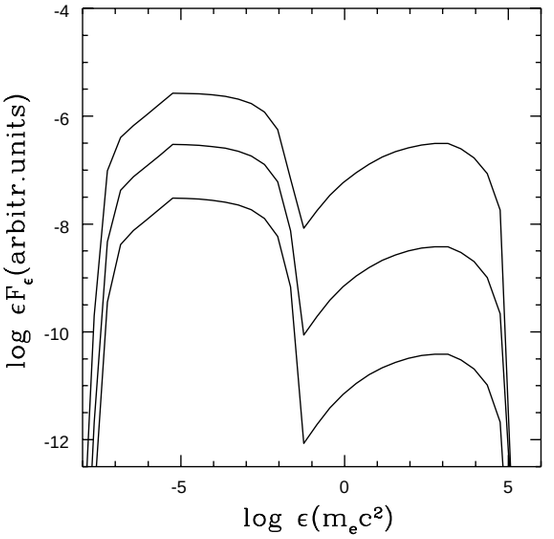

Fig. 7. Steady-state synchrotron and SSC emission produced from a power law electron distribution function for different normalizations of the electrons. As this changes, it produces linear changes to the synchrotron component (left 'hump') and quadratic changes to the SSC component (right 'hump').

between $\epsilon_{ssc,min} \simeq \epsilon_{s,min}\gamma_{min}^2 \simeq b\gamma_{min}^4$ and $\epsilon_{ssc,max} \simeq \epsilon_{s,max}\gamma_{max}^2 \simeq b\gamma_{max}^4$, therefore it will be twice more spread, in logarithmic scale, than the synchrotron component. However the two components are expected to have the same slope as Thomson-limit ICS spectra have the same slope as the synchrotron spectra when coming from the same electron population (see previous section).

Complications in this simple picture are introduced from the fact that in many cases collisions between the high energy tail of the synchrotron emission and the high end of electron distribution are in the Klein-Nishina regime. In this case simple estimates are difficult and numerical calculations become necessary.

8 Continuity Equation

In the previous sections we showed that the emitted synchrotron and inverse Compton spectra depend on the underlying electron spectrum. On the other hand, both processes act as energy loss mechanisms on the relativistic electrons and therefore they play an important role in the formation of the electron distribution function. The fact that high energy electrons lose their energies faster than the lower energy ones (see Eqn 30) leads naturally to the problem of particle evolution and aging within the sources. This becomes all the more relevant from observations of miniquasars, AGNs and GRBs that show intense flux variations, implying that a steady-state power law of electrons cannot treat the problem self-consistently.

A related problem arises from the realization that different energy loss mechanisms could operate on the same particles. These will leave, in general, different signatures on the particle distribution and therefore on the radiated photon spectra. For example, if high energy electrons are injected in some volume which is expanding, then one has to consider adiabatic losses in addition to synchrotron and inverse Compton scattering losses. If, furthermore, the number density of ambient electrons and protons is high, then bremsstrahlung losses might play an important role and so on. Therefore to calculate the distribution function of the electrons one has to carefully identify, as a first step, the possible energy loss mechanisms acting on them. In addition the question of the high energy electron sources has to be addressed. This question is much more difficult to answer and it invites a close examination of the acceleration mechanisms (see Gallant, these Proceedings).

Therefore, a simple way to model the emission from time variable synchrotron/inverse Compton sources is to first calculate the electron distribution function. To do so one has to solve the so-called *continuity* equation which follows the evolution of particles in the source by including, as mentioned above, basically particle energy loss mechanisms and particle injection (or acceleration) terms. Historically this type of equation has been used to describe the evolution of cosmic rays as they diffuse in the Galaxy (see, for example, [7]). In our case the prescription is slightly different as the particles are assumed to be injected uniformly within some spherical source of radius R. Then the continuity equation can be written in a most general form as

$$\frac{\partial n_e(\gamma,t)}{\partial t} = \mathcal{Q}^e + \mathcal{L}^e \tag{54}$$

where $\mathcal{Q}^e$ represent the rates of the sources of relativistic electrons, while $\mathcal{L}^e$ represent the rates of various energy (and particle) loss mechanisms. One can complicate this equation according to need (or taste) by including various relevant processes. For example, as sources of high energy electrons one could use, apart from the already mentioned acceleration, electron-positron pair injection either from photon-photon pair production or from hadron-photon interactions. On the other hand, as electron energy or particle loss mechanisms one could use synchrotron and inverse Compton scattering in the former and electron escape or electron-positron annihilation in the latter case.

The equation above should be complemented with a corresponding kinetic equation for photons

$$\frac{\partial n_\gamma(\epsilon,t)}{\partial t} + \frac{n_\gamma(\epsilon,t)}{t_{esc}} = \mathcal{Q}^\gamma + \mathcal{L}^\gamma \tag{55}$$

where, in complete analogy to the previous equation, $\mathcal{Q}^\gamma$ denote the rates of the photon sources (for example, synchrotron and inverse Compton emissivities), while $\mathcal{L}^\gamma$ are the rates for photon sinks (for example, photon-photon pair production). The second term on the left hand side of (55) introduces the escape time from the source which is set equal to R/c implying that the radius of the

source R is a free boundary– this is a necessary simplification to the radiation transfer problem.

Eqns (54) and (55) are coupled and, according to the processes chosen, they can be non-linear. There are many versions of such type of equations in the astrophysical literature, mostly developed for the description of optically thick electron-positron pair plasmas, either in steady state (see, e.g. [18]) or time-dependent [5]. Many examples of the time evolution of the electrons for various types of losses and source terms are given in [16]. To keep a continuity with the previous sections we will present first an example of how Eqns (54) and (55) can be solved in the case of electron power law injection and synchrotron cooling in an approach borrowed from [16] – and revisited in [21].

We assume therefore that at a time $t = 0$ a power law of electrons between two limiting values γ_{min} and γ_{max} is uniformly injected within some spherical source of radius R. Since (again by assumption) synchrotron is the dominant loss mechanism operating on the electrons we can write from equation (54)

$$\frac{\partial n_e}{dt} + \frac{4}{3}\sigma_T c \left(\frac{U_B}{m_e c^2}\right) \frac{\partial}{\partial \gamma}\left(\gamma^2 n_e\right) = Q_0 \gamma^{-s} \tag{56}$$

where Q_0 is the rate with which particles are injected into the source. The second term on the left hand side of (56) is the synchrotron loss term. Using as time unit the source crossing time $t_c = R/c$ and introducing the 'magnetic compactness'

$$l_b = \sigma_T R \left(\frac{U_B}{m_e c^2}\right) \tag{57}$$

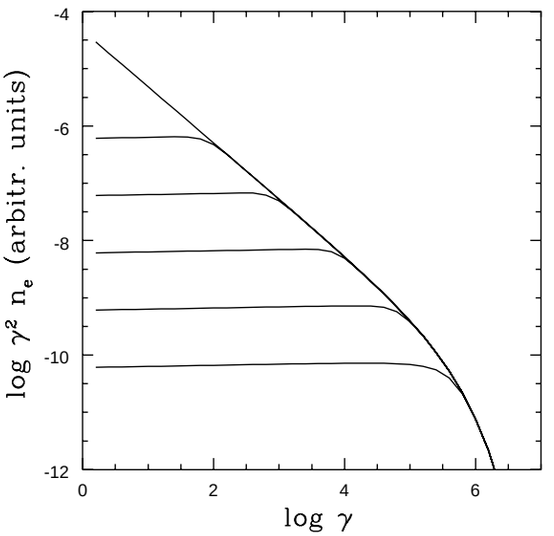

Fig. 8. Snapshots of the electron distribution function for constant amplitude power-law electron injection and synchrotron cooling. The slope of injected electrons was taken to be $s = 2$. The system evolves from bottom to top.

we can rewrite equation (56)

$$\frac{\partial n_e}{dt} + \frac{4}{3} l_B \frac{\partial}{\partial \gamma} \left(\gamma^2 n_e\right) = \hat{Q}_0 \gamma^{-s} \tag{58}$$

where $\hat{Q}_0 = Q_0 R/c$ and the time t is measured in units of R/c.

The above equation can be solved analytically (see [16]) and then the radiated spectrum can be obtained by folding the electron distribution with the single particle synchrotron emissivity, as was shown in Section 5. In Figs. 8 and 9 we show respectively the evolution of the electron distribution and of the photon spectrum for various instances after electrons start getting injected and radiating in the source. From the characteristic equation one can calculate the break energy

$$\gamma_{br}(t) = \gamma_{max}/(1 + 4\gamma_{max} l_B t/3) \tag{59}$$

which denotes the energy above which the distribution had had time to cool. Indeed for $\gamma_{br} < \gamma < \gamma_{max}$ the electron distribution is given by

$$n_e(\gamma, t) = \frac{3\hat{Q}_0}{4 l_B (s-1) \gamma^2} \left[\gamma^{-(s-1)} - \gamma_{max}^{-(s-1)}\right] \tag{60}$$

which for $\gamma \ll \gamma_{max}$ gives $n_e \propto \gamma^{-(s+1)}$, i.e. the spectrum has steepened by in comparison to the injected one. On the other hand, for $\gamma < \gamma_{br}$ the electron spectrum had no time to cool and it is given by

$$n_e(\gamma, t) = \frac{3\hat{Q}_0 \gamma^{-s}}{4 l_B (s-1)} \left[\gamma^{-1} - \gamma^{-1}(1 - \frac{4}{3} l_B \gamma t)^{s-1}\right]. \tag{61}$$

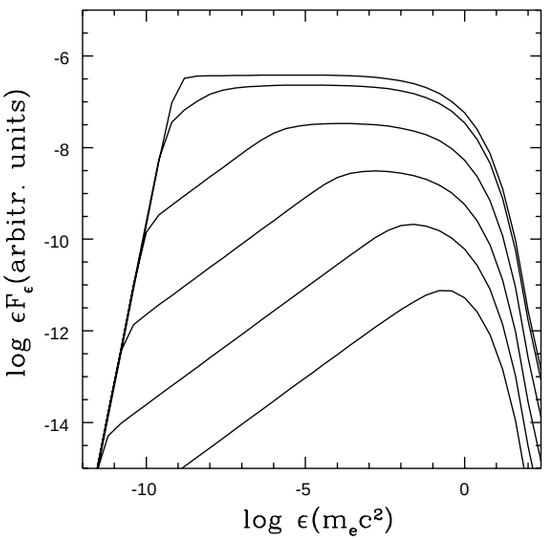

Fig. 9. The synchrotron photon spectra corresponding to the electron distribution snapshots of the previous figure.

which for $4l_B\gamma t/3 \ll 1$ gives $n_e \simeq \hat{Q}_0 t\gamma^{-s}$, i.e. the index is the same as the one at injection, while the number of electrons at the particular energies increases linearly with time. These traits are evident in Fig. 8 which shows snapshots of the distribution function at times $t_1,\ t_2, ...,\ t_6$ such that $\gamma_{max} = 10\gamma_{br}(t_1) = 10^2\gamma_{br}(t_2)$ and so on. Only the last time t_6 was chosen as to allow the distribution to reach a steady state.

Fig. 9 shows the corresponding photon spectra. The break at low energies is due to synchrotron self absorption and can be obtained by introducing a suitable absorption rate at Eqn (55). The spectrum below that break, which we will assume that occurs at an energy ϵ_{ssa}, has a flat spectral index of 5/2 irrespective of the slope of the electron distribution. At higher energies there is a second spectral break at some energy ϵ_{br} which corresponds to the electron break at γ_{br} and is related with it through the relation $\epsilon_{br} = b\gamma_{br}^2$. We note that since γ_{br} has a time dependence (c.f. Eqn 59), we expect ϵ_{br} to be dependent also on time and this is clearly seen in Fig. 9. According to synchrotron theory (see Section 5) the spectral index a of the photons is related to the particle index p by $a = (p-1)/2$. Then from Eqn (61) we get that $a = (s-1)/2$ for energies less than ϵ_{br}, while Eqn (60) gives $a = s/2$ for energies larger than ϵ_{br}, i.e. the photon spectrum steepens by 1/2, which is one of the characteristic features of synchrotron radiation[4]. Therefore we can summarise the above results as follows:

$$\begin{aligned} &\text{For} \quad \epsilon \ll \epsilon_{ssa} && \rightarrow a = 5/2 \\ &\text{For} \quad \epsilon_{ssa} \ll \epsilon \ll \epsilon_{br} && \rightarrow a = (s-1)/2 \\ &\text{For} \quad \epsilon_{br} \ll \epsilon \ll \epsilon_{max} && \rightarrow a = s/2 \end{aligned}$$

As another example we show the opposite case, i.e. electrons stop being injected into a source after some time t_0. Fig. 10 shows the photon spectrum evolution for times $t > t_0$ with time increasing as the spectra evolve from high to low energies (right to left). As in the previous example the times at which these snapshots were taken are equally spaced in logarithm. It is seen that as time increases first the higher energy and then the lower photons start getting depleted. This is because high energy electrons cool faster, so they will be the first to lose their energies and as there is no replenishment of fresh electrons the photon emission around the critical frequency of these electrons will drop. At later times even less energetic electrons will have time to cool and so on.

The final example (Fig. 11) shows a fully materialized application of the kinetic equation approach to model the flaring activity of the TeV blazars [22]. Electrons are continuously injected into and cool in a blob moving with a Doppler factor δ. The equations are solved in the rest frame of the blob and the results are then transformed to the observer's frame as explained in section 3. The processes include, apart from the already mentioned electron injection, synchrotron radiation, inverse Compton scattering in both Thomson and Klein Nishina regimes and electron escape. Other processes like photon-photon pair production or

[4] Obviously the same results are obtained for inverse Compton scattering assuming that all interactions are in the Thomson regime

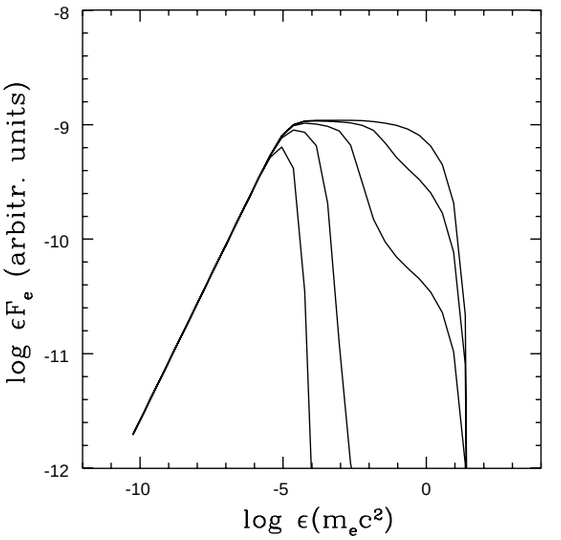

Fig. 10. Snapshots of synchrotron photon spectra resulting after electrons stop being injected. The photons are depleted progressively from high to low energies, which reflects the corresponding electron cooling.

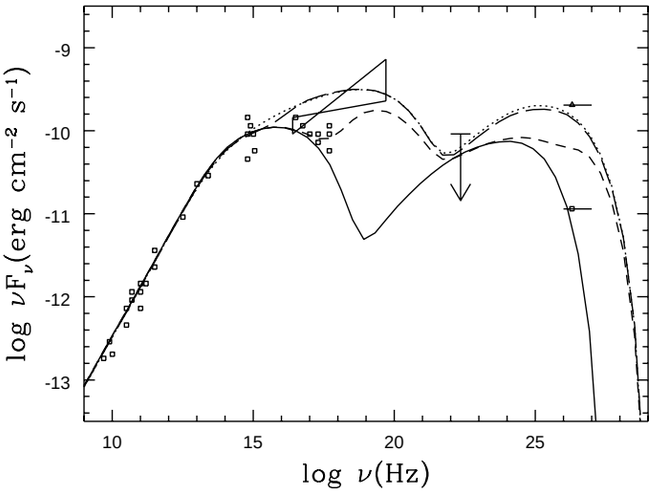

Fig. 11. A fit to a X-ray/TeV $\gamma-$ray flare of Mkn501. Injected electrons are assumed to increase impulsively their maximum energy and this produces a short (timescale of hours) flare - dashed lines, until the new steady state is reached - dotted line.

bremsstrahlung turn out to be marginal (for the chosen parameters) and can safely be neglected. In these models radio to X-rays are produced from synchrotron radiation of one population of electrons, while low energy (MeV) to

very high energies (TeV) are produced from the synchro self-Compton mechanism. Flares can be generated by changing one (or more) parameter of the injected electrons and then followed in a time dependent fashion by using equations (54) and (55). Such a flare is shown in Figure (11) where the maximum cutoff of the electron distribution has been increased impulsively. This change produces a flaring activity at the high energy cutoff of the photon distribution, i.e. in the X and TeV regime, and is a way of explaining the recent blazar multiwavelength obervations – [12], [22], [20], [17]. Similar methods have been applied to GRBs – [6], [24].

9 Coda

In the previous sections we gave short reviews of some key phenomena and processes that are related with objects showing relativistic outflows. The physical processes described above assume that relativistic *electrons* are responsible for the observed non-thermal radiation from sources such as blazars and GRBs. However, it is not clear at this stage whether this is the whole story or whether there is, as well, a relativistic *hadronic* component present in these sources. This latter possibility introduces a new dimension both to the theoretical and to the observational aspects of these objects. In modeling one has to take into account processes such as proton-proton and proton-photon interactions. These act both as proton energy loss and as injection mechanisms for secondary products. The secondaries include very high energy $\gamma-$rays, electron-positron pairs, neutrons and neutrinos[5]. If we would have chosen the kinetic equation approach to describe the system (see Section 8) we should have added an extra equation, i.e. the one for protons, in addition to the ones for electrons and photons. To do so suitable expressions for the various proton rates had to be chosen (see [1]) and in order to follow the evolution of the system we would have to solve the three resulting coupled partial differential equations (for examples see [21]). On the other hand, the (necessary) approximations introduced with the use of the various production rates (exclusive use of the Δ resonance for the cross section calculation, multiplicities, etc) can be alleviated with the use of Monte-Carlo codes such as the recently built SOPHIA (see, for example, [23]) which, however, still requires a full implementation to astrophysical sources. Concluding we should stress that the possibility of the presence of relativistic hadrons in astrophysical sources like AGNs and GRBs remains an open and exciting one with far-fetching consequences for Cosmic Ray Physics and Neutrino Astronomy.

Acknowledgments: I would like to thank the Organizing Committee for organizing a very successful, enjoyable and well planned down to the last detail Conference.

[5] Especially neutrino emission is a unique feature of the hadronic models and maybe it is the only way to prove (or disprove) the existence of relativistic hadrons in AGNs and GRBs

References

1. M.C. Begelman, B. Rudak, M. Sikora: Ap.J. **362**, 38, (1990)
2. R.D. Blandford, H. Netzer, L. Woltjer: *Active Galactic Nuclei*, Saas-Fe Advanced Course, (Springer-Verlag 1990)
3. G.R. Blumenthal, R.J. Gould: Rev.Mod.Phys. **42**, 237 (1970)
4. M.H. Cohen, W. Cannon, et al. : Ap.J. **170**, 207 (1971)
5. P.S. Coppi: M.N.R.A.S. **258**, 657 (1992)
6. C.D. Dermer, J. Chiang, K.E. Mitman: Ap.J. **537**, 785 (2000)
7. V.L. Ginzburg, S.I. Syrovatskii: *The Origin of Cosmic Rays*, (Pergamon Press 1964)
8. V.L. Ginzburg, S.I. Syrovatskii: A.R.A.& A. **3**, 297 (1965)
9. R.J. Gould, G.P. Schreder: Phys.Rev. **155**, 1408 (1967)
10. P.W. Guilbert, A.W. Fabian, M.J. Rees: M.N.R.A.S. **205**, 593 (1983)
11. F. Hoyle: M.N.R.A.S. **120**, 338 (1960)
12. S. Inoue, F. Takahara: Ap.J. **463**, 555 (1996)
13. J.D. Jackson: *Classical Electrodynamics* (J. Wiley and Sons, New York 1975)
14. F.C. Jones: Phys.Rev. **167**, 1159 (1968)
15. T.W. Jones, S.F. O'Dell, W.A. Stein: ApJ **188**, 353 (1974)
16. N.S. Kardashev: Sov.Astr.J. **6**, 317 (1962)
17. H. Krawczynski, P.S. Coppi, T. Maccarone, F.A. Aharonian: A.&A. **353**, 97 (2000)
18. A.P. Lightman, A.A. Zdziarski: Ap.J. **253**, 842 (1987)
19. K.R. Lind, R.D. Blandford : Ap.J. **295**, 358 (1985)
20. L. Maraschi, G. Fossati, F. Tavecchio, et al.: Ap.J. **526**, 81 (1999)
21. A. Mastichiadis, J.G. Kirk: A.& A. **295**, 613 (1995)
22. A. Mastichiadis, J.G. Kirk: A.& A. **320**, 19 (1997)
23. A. Mucke, J.P. Rachen, R. Engel, R.J. Protheroe, T. Stanev: P.A.S.A. **16(2)**, 160 (1999)
24. A. Panaitescu, P. Meszaros: Ap.J. **544**, L17 (2000)
25. M.J. Rees: Nature **211**, 468 (1966)
26. G.B. Rybicki, A.P. Lightman: *Radiative Processes in Astrophysics* (Wiley Interscience, New York 1979)

Particle Acceleration at Relativistic Shocks

Yves A. Gallant[1,2]

[1] Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
[2] Astronomical Institute Utrecht, Postbus 80 000, 3508 TA Utrecht, Netherlands

Abstract. I review the current status of Fermi acceleration theory at relativistic shocks. I first discuss the relativistic shock jump conditions, then describe the non-relativistic Fermi mechanism and the differences introduced by relativistic flows. I present numerical calculations of the accelerated particle spectrum, and examine the maximum energy attainable by this process. I briefly consider the minimum energy for Fermi acceleration, and a possible electron pre-acceleration mechanism.

1 Introduction and Motivation

A ubiquitous feature of astrophysical objects involving relativistic flows, such as active galactic nuclei (AGNs), gamma-ray bursts (GRBs) and Crab-like supernova remnants (SNRs), is the presence of nonthermal, power-law emission spectra (i.e. with flux density $F_\nu \propto \nu^{-\alpha}$, where ν is the frequency and α the spectral index), in particular in the radio and hard X-ray or gamma-ray domains. This emission is believed to be produced by accelerated particles having a corresponding power-law energy spectrum; more specifically, in most of these objects the emission is thought to be from accelerated electrons radiating via the synchrotron or inverse Compton mechanisms (see Mastichiadis, this volume). The aim of the present review will be to discuss the probable mechanism of this acceleration and the spectra that may be expected theoretically.

The most widely invoked mechanism for the acceleration of particles to power-law spectra in non-relativistic contexts, such as SNR blast waves or interplanetary shocks, is *Fermi acceleration*. It seems likely that shocks are responsible for particle acceleration in relativistic flows as well, and this is indeed explicitly assumed in models of GRBs and Crab-like SNRs. It is then natural to consider how the Fermi mechanism could operate at relativistic shocks, and what the resulting spectrum would be. The focus of this contribution will thus be the relativistic version of Fermi shock acceleration.

This review is organised as follows: in Sect. 2, I discuss the shock jump conditions at relativistic shocks, emphasising the aspects relevant to particle acceleration; this section is intended to be self-contained. In Sect. 3, I describe the Fermi acceleration mechanism in detail, first reviewing its main features in the context of non-relativistic shocks, and then presenting the resulting spectrum for ultra-relativistic and more moderately relativistic shocks. In Sect. 4, I examine the acceleration time scale and the maximum energy attainable by this mechanism, and consider the minimum energy for Fermi acceleration of electrons in an electron–ion shock, and a possible pre-acceleration mechanism.

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 24–40, 2002.

2 Relativistic Shocks

The properties of shocks most important for Fermi-type particle acceleration are the velocities of the shock relative to the upstream and downstream frames. These are obtained through the *shock jump conditions*.

2.1 Relativistic Shock Jump Conditions

The shock jump conditions are derived from the laws of conservation of particle number, energy, and momentum. For relativistic fluids, these are, in order:

$$\Gamma_1\beta_1 n_1 = \Gamma_2\beta_2 n_2 \ , \tag{1}$$

$$\Gamma_1^2\beta_1(\varepsilon_1 + p_1) = \Gamma_2^2\beta_2(\varepsilon_2 + p_2) \ , \tag{2}$$

$$\Gamma_1^2\beta_1^2(\varepsilon_1 + p_1) + p_1 = \Gamma_2^2\beta_2^2(\varepsilon_2 + p_2) + p_2 \ , \tag{3}$$

where subscripts 1 and 2 respectively refer to the upstream and downstream regions, n, ε and p are the fluid number density, energy density and pressure, all measured in the local fluid rest frame, β is the fluid velocity in units of the speed of light c, and Γ the corresponding Lorentz factor. The fluid velocities are measured in the *shock frame*, where the shock is stationary and both velocity vectors lie along the shock normal. Equivalently, β_1 and β_2 may be viewed as the shock velocity with respect to the upstream and downstream fluids.

For simplicity, I restrict my attention in this review to unmagnetised shocks. The shock jump conditions for relativistic magneto-hydrodynamics (MHD) were reviewed by Kirk and Duffy [20]. They can yield shock jump conditions which differ significantly from those derived here when the magnetisation parameter,

$$\sigma \equiv \frac{B_1^2}{4\pi(\varepsilon_1 + p_1)} \ , \tag{4}$$

where B is the magnetic field measured in the local fluid rest frame, is not negligibly smaller than unity [19,21].

2.2 Ultra-relativistic Shocks

Much of the discussion of particle acceleration below will be specialised to ultra-relativistic shocks, i.e. those in which the shock Lorentz factor $\Gamma_{\rm sh} \equiv \Gamma_1 \gg 1$, and $\beta_1 \approx 1$. In that case the pressure term p_1 may be neglected in (3) relative to the first term. The particles downstream of such a shock must be heated to highly relativistic temperatures; assuming they obey the ultra-relativistic gas equation of state, $\varepsilon_2 = 3p_2$, one may then solve (2) and (3) to obtain the *ultra-relativistic shock jump conditions*, yielding the downstream velocity

$$\beta_2 \approx \frac{1}{3} \ . \tag{5}$$

Also of interest is the relative velocity $\beta_{\rm rel}$ of the upstream and downstream fluids. Using the relativistic velocity addition formula,

$$\beta_{\rm rel} = \frac{\beta_1 - \beta_2}{1 - \beta_1 \beta_2} \, , \tag{6}$$

the associated Lorentz factor in the ultra-relativistic limit is $\Gamma_{\rm rel} \approx \Gamma_{\rm sh}/\sqrt{2}$. It should be noted that these shock jump conditions are independent of the upstream equation of state, and depend only on the downstream gas being ultra-relativistically hot. It will be seen below that particle acceleration in the ultra-relativistic shock regime thus mirrors some of the simplicity of the non-relativistic, strong shock regime, due to the existence of this single, well-defined asymptotic value (for weakly magnetised shocks) of the shock velocity ratio.

2.3 Moderately Relativistic Shocks

For more general values of the shock Lorentz factor $\Gamma_{\rm sh}$, the shock jump conditions depend on the equation of state and the temperature of the upstream gas. For illustration, the two opposite extremes of an ultra-relativistically hot and a cold gas upstream will be examined.

Shocks in an Ultra-relativistic Gas

I first assume that the gas upstream of the shock already has a highly relativistic temperature; this might be the case for an internal shock in a GRB fireball, for instance, if it propagates in a medium already heated by previous shell collisions (see e.g. Sari & Galama, this volume). In this case, both the upstream and the downstream media can be assumed to follow the ultra-relativistic equation of state, $\varepsilon = 3p$. Equations (2) and (3) are then readily solved to yield the jump condition for shocks propagating in an ultra-relativistic gas:

$$\beta_1 \beta_2 = \frac{1}{3} \, . \tag{7}$$

This relation holds for shocks of any strength, provided only that the upstream gas is ultra-relativistic. The only requirement for the existence such of a shock solution is that the upstream flow velocity be larger than the sound speed in the upstream gas, which is $c/\sqrt{3}$ for an ultra-relativistic gas.

Strong Shocks and the Synge Equation of State

I now consider the case of a *strong shock*, i.e. one in which the thermal energy upstream is negligible with respect to the bulk flow kinetic energy, so that one may neglect the upstream pressure p_1 and write $\varepsilon_1 \approx n_1 mc^2$ in (2) and (3), where m is the mass of individual gas particles, assumed for simplicity to belong

to a single species. For the downstream equation of state, I use that of an ideal gas of arbitrary temperature, as given by Synge [29]:

$$\varepsilon_2 + p_2 = n_2 mc^2 G\left(\frac{mc^2}{T_2}\right) . \tag{8}$$

Here T_2 is the downstream gas temperature, and the function $G(\xi)$ is defined in terms of modified Bessel functions of the first kind, $G(\xi) \equiv K_3(\xi)/K_2(\xi)$, and has the asymptotic expansions:

$$G\left(\frac{mc^2}{T}\right) = 1 + \frac{5}{2}\frac{T}{mc^2} + \mathcal{O}\left(\frac{T}{mc^2}\right)^2 , \qquad T \ll mc^2 , \tag{9}$$

$$G\left(\frac{mc^2}{T}\right) = \frac{4T}{mc^2} + \frac{mc^2}{2T} + \mathcal{O}\left(\frac{mc^2}{T}\right)^3 , \qquad T \gg mc^2 . \tag{10}$$

Using the fact that the gas always obeys the ideal gas law $p = nT$, it is readily seen that these two asymptotes correspond to the familiar equations of state for non-relativistic and ultra-relativistic ideal gases, namely $\varepsilon = nmc^2 + 3p/2$ and $\varepsilon = 3p$, respectively.

With the above assumptions, the shock jump conditions (1–3) may be solved by first using (1) to rewrite (2) and (3) in terms of the normalised quantities

$$\bar{\varepsilon}_2 \equiv \frac{\varepsilon_2}{n_2 mc^2} = G(\xi) - \frac{1}{\xi} , \tag{11}$$

$$\bar{p}_2 \equiv \frac{p_2}{n_2 mc^2} = \frac{1}{\xi} , \tag{12}$$

where the *reciprocal temperature* of the downstream gas has been defined as $\xi \equiv mc^2/T_2$. The two resulting equations may be solved to yield Γ_1 and Γ_2 in terms of ξ [20]:

$$\Gamma_2^2 = \frac{\bar{\varepsilon}_2^2 - 1}{\bar{\varepsilon}_2^2 - \bar{p}_2^2 - 1} , \tag{13}$$

$$\Gamma_1 = (\bar{\varepsilon}_2 + \bar{p}_2)\Gamma_2 . \tag{14}$$

If one prefers to use Γ_1 rather than T_2 as the independent variable, the analytical equation (14), substituting the definitions (11–13), may be inverted numerically. The shock velocity ratio β_1/β_2 resulting from (13–14) is shown in Fig. 1 as a function of the upstream four-velocity $\beta_1\Gamma_1$, along with the corresponding result for the ultra-relativistic upstream gas case (7).

Electron–Ion Plasmas: The shock jump conditions derived above, based on the equation of state (8), are strictly speaking only valid for a gas composed of particles of a single mass m, as would be the case for instance in an electron–positron plasma. Equation (8) is readily generalised to a gas composed of species of different masses [20,29], assuming they are in thermal equilibrium. However,

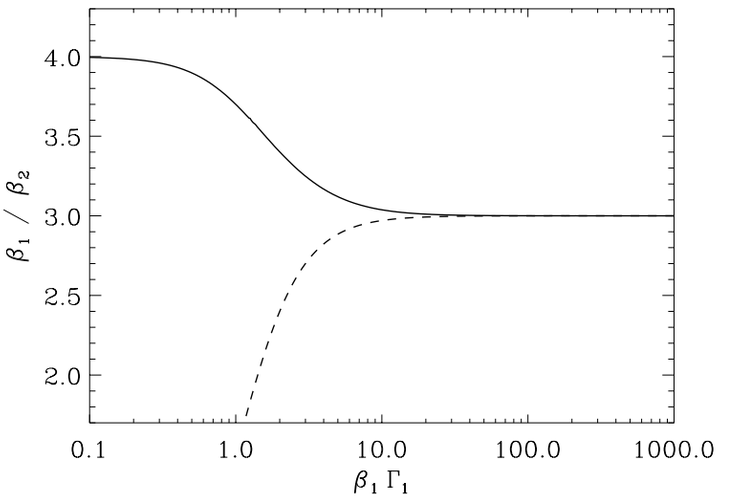

Fig. 1. The shock velocity ratio as a function of the upstream four-velocity, for the two extreme cases of a cold gas upstream obeying the Synge equation of state downstream (*solid line*), and of an ultra-relativistically hot gas upstream (*dashed line*).

simulations of relativistic, perpendicular shocks in electron–ion plasmas [18] show that the species do not in fact achieve thermal equilibrium immediately behind the shock, but instead have distinct temperatures corresponding to the thermalisation of their respective upstream bulk kinetic energy. The strong shock jump condition illustrated in Fig. 1 then also applies in this case, as I now demonstrate.

As remarked in [20], for a strong shock one may derive from the shock jump conditions (1–3) the relation

$$\varepsilon_2 = \Gamma_{\rm rel} \rho_2 c^2 \ , \tag{15}$$

where ρ_2 is the total downstream rest mass density, and $\Gamma_{\rm rel}$ the Lorentz factor corresponding to the relative velocity (6). This equation shows that the downstream energy density per unit mass is simply the upstream bulk flow energy of the particles as seen from the downstream frame. If this relation holds separately for each species, as is the case initially in the electron–ion shock simulations mentioned above, one has for each species, using (11):

$$G(\xi) - \frac{1}{\xi} = \Gamma_{\rm rel} \ . \tag{16}$$

Thus although the temperatures of the species will in general be different, their normalised (reciprocal) temperatures ξ will be the same, and the shock jump conditions obtained above for a single particle mass will hold in this case also.

Energy exchange between the electrons and the ions does take place downstream of the shock in the above-mentioned simulations, but appears to result

in a power-law tail of the electron energy distribution rather than simple heating, as discussed further in Sect. 4.4. One could envision shock jump conditions taking into account this phenomenon in the downstream equation of state; however, an equally important component of more realistic shock jump conditions is the energy and momentum carried away by the strong electromagnetic precursor emitted by the shock front [16]. As neither of these two phenomena can be predicted quantitatively at present, it seems premature to attempt to obtain more accurate shock jump conditions for electron–ion plasmas than that shown in Fig. 1 for a simple Synge equation of state.

3 Fermi Acceleration and the Spectral Index

In this section, I first review the basic ideas of the Fermi acceleration mechanism in the context of non-relativistic shocks, then discuss in some detail its application to the opposite extreme of ultra-relativistic shocks and the spectral index resulting in that case, before addressing the more involved intermediate case of moderately relativistic shocks.

3.1 Non-relativistic Shocks

While the essential concepts of the acceleration mechanism that bears his name date back to Fermi [10], their application to shock acceleration was first proposed in 1977–78 in four independent papers [2,4,9,23]; of these, I will follow most closely below the treatment given by Bell [4].

The acceleration scenario is illustrated schematically in Fig. 2: a high-energy particle, assumed for simplicity to be already relativistic, diffuses through the medium on either side of the shock by scattering on magnetic irregularities. These may be, for instance, Alfvén waves self-consistently excited by the diffusing high-energy particles [5]. Assuming the local Alfvén velocity is much smaller than the shock velocity, to lowest order the magnetic scattering centres may be considered at rest with respect to the fluid, so that the scattering events do not change the particle energy in the local fluid rest frame.

Consider, then, a particle diffusing upstream or downstream while preserving its energy in the corresponding rest frame. A particle initially having energy E_i in the upstream medium will eventually cross the shock, its velocity upon crossing making an angle $\theta_{\to\mathrm{d}}$ with the shock normal (see Fig. 2). Its energy measured in the downstream frame, E'_i, will then be given by the appropriate Lorentz transformation, and preserved while the particle is downstream. If it re-crosses the shock into the upstream medium, this time at an angle $\theta'_{\to\mathrm{u}}$, its final energy upstream, E_f, will be given by the combination of the two Lorentz transformations:

$$\frac{E_\mathrm{f}}{E_\mathrm{i}} = \Gamma_\mathrm{rel}^2(1-\beta_\mathrm{rel}\mu_{\to\mathrm{d}})(1+\beta_\mathrm{rel}\mu'_{\to\mathrm{u}}) \;, \tag{17}$$

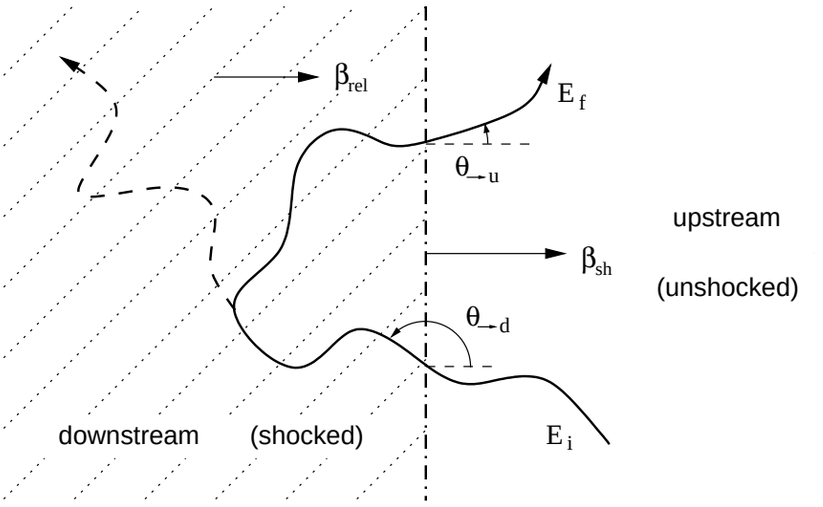

Fig. 2. Schematic representation of one cycle of shock acceleration: a shock propagates with velocity $\beta_{\rm sh}$ into the undisturbed (upstream) medium to the right; the velocity of the shocked (downstream) medium relative to the upstream one is $\beta_{\rm rel}$. A relativistic particle diffuses through the media on both sides of the shock, crossing and re-crossing it at incidence angles $\theta_{\rightarrow \rm d}$ and $\theta_{\rightarrow \rm u}$, with initial and final energies $E_{\rm i}$ and $E_{\rm f}$.

where $\beta_{\rm rel}$ and $\Gamma_{\rm rel}$ are the relative velocity of the upstream and downstream media and the corresponding Lorentz factor, and I have introduced the notation μ for $\cos\theta$. Here and in what follows primed and unprimed quantities are respectively measured in the downstream and upstream rest frames. The only approximation made in deriving (17) is that that the particle is highly relativistic, so that the rest mass contribution to its energy may be neglected.

For non-relativistic shocks, $\beta_{\rm rel} \ll 1$, the angular distribution of these scattered particles crossing the shock may be approximated as isotropic, so that the flux-weighted averages of the direction angle cosines, over the relevant ranges $-\pi/2 \leq \theta_{\rightarrow \rm d} \lesssim 0$ and $0 \lesssim \theta'_{\rightarrow \rm u} \leq \pi/2$, are respectively $\langle \mu_{\rightarrow \rm d} \rangle \approx -2/3$ and $\langle \mu'_{\rightarrow \rm u} \rangle \approx 2/3$. The average energy gain (17) per shock crossing cycle then reduces to

$$\left\langle \frac{E_{\rm f}}{E_{\rm i}} \right\rangle \approx 1 + \beta_{\rm rel} \left(\langle \mu'_{\rightarrow \rm u} \rangle - \langle \mu_{\rightarrow \rm d} \rangle \right) \approx 1 + \frac{4}{3} \beta_{\rm rel} \ . \tag{18}$$

Fermi shock acceleration is sometimes referred to as the *Fermi I* mechanism because this energy gain is of first order in the velocity $\beta_{\rm rel}$.

While a particle upstream will always eventually cross the shock, at least in the simple case considered here of an infinite, plane-parallel shock, once downstream the particle has a certain probability of being advected away and never re-crossing the shock. This *escape probability* may be evaluated from the ratio of the average fluxes of particles escaping far downstream and crossing the shock.

For a downstream flow velocity β_2, again assuming isotropy of the relativistic particle distribution, it is given by $P_{\rm esc} = 4\beta_2$.

The combination of the energy gain factor (18) and escape probability $P_{\rm esc}$ leads to a power-law spectrum in the accelerated particle energy, with a spectral index depending solely on the shock jump conditions:

$$\frac{{\rm d}N}{{\rm d}E} \propto E^{-(r+2)/(r-1)} \ , \tag{19}$$

where $r \equiv \beta_1/\beta_2$ is the shock *velocity ratio*. For a strong, non-relativistic shock in a monatomic gas, $r = 4$, leading to a spectral index $(r+2)/(r-1) = 2$ for the accelerated particles. In reality, spectral indices somewhat steeper than this value are often observed; this difference may be due to the pressure of the accelerated particles modifying the shock structure (e.g. [7] and references therein).

3.2 Ultra-relativistic Shocks

I now turn my attention to ultra-relativistic shocks, i.e. those for which $\Gamma_{\rm sh} \gg 1$ so that the shock jump conditions derived in Sect. 2.2 apply.

Energy Gain and Upstream Particle Dynamics

For a downstream particle to cross the shock into the upstream medium, it must have $1 \geq \mu'_{\rightarrow {\rm u}} > \beta'_{\rm sh} = \frac{1}{3}$, so that the factor $(1 + \beta_{\rm rel}\mu'_{\rightarrow {\rm u}})$ in (17) is always of order unity. If $\mu_{\rightarrow {\rm d}}$ is approximately isotropically distributed, as might be the case for a population of relativistic particles already present in the undisturbed upstream medium, the factor $(1-\beta_{\rm rel}\mu_{\rightarrow {\rm d}})$ is in general also of order unity. Thus in the first shock crossing cycle, a large *initial boost* in energy can be achieved, $E_{\rm i}/E_{\rm f} \sim \Gamma_{\rm rel}^2$ as envisioned in [30].

For all subsequent shock crossing cycles, however, the distribution of $\mu_{\rightarrow {\rm d}}$ will be highly anisotropic; this is an essential difference between non-relativistic and relativistic Fermi shock acceleration [22,27]. For an ultra-relativistic particle with Lorentz factor $\gamma \gg \Gamma_{\rm sh}$, the kinematic condition to cross the shock into the upstream medium reduces to $\theta_{\rightarrow {\rm u}} < 1/\Gamma_{\rm sh}$. As shown in [13], for realistic deflection processes upstream the particle cannot be deflected very far beyond this 'loss cone' before the shock overtakes it, so that $\theta_{\rightarrow {\rm d}} \sim 1/\Gamma_{\rm sh}$ as well. In this case the energy gain factor reduces to

$$\frac{E'_{\rm f}}{E'_{\rm i}} \approx \frac{2 + (\Gamma_{\rm sh}\theta_{\rightarrow {\rm d}})^2}{2 + (\Gamma_{\rm sh}\theta_{\rightarrow {\rm u}})^2} \approx \frac{1 + \mu'_{\rightarrow {\rm u}}}{1 + \mu'_{\rightarrow {\rm d}}} \ , \tag{20}$$

where the shock crossing cycle is now considered from downstream to upstream and back.

The range of possible energy gain factors can be assessed by considering two opposite extremes for the upstream particle dynamics: deflection by a regular magnetic field and scattering by small-scale magnetic fluctuations. In terms of the correlation length ℓ of the magnetic field, these two regimes respectively

correspond to $R_{\rm L}/\Gamma_{\rm sh} \ll \ell$ and $R_{\rm L}/\Gamma_{\rm sh} \gg \ell$, where $R_{\rm L}$ is the Larmor radius of the particle. For regular deflection, it can be shown that for ingress angles $0 \leq \Gamma_{\rm sh}\theta_{\rightarrow {\rm u}} < 1$, the egress angle satisfies

$$1 < \Gamma_{\rm sh}\theta_{\rightarrow {\rm d}} \leq 2 \qquad \Longleftrightarrow \qquad \frac{1}{3} > \mu'_{\rightarrow {\rm d}} \geq -\frac{1}{3}\,, \tag{21}$$

while for direction-angle scattering, the direction angle at the average shock recrossing time satisfies

$$\langle \theta^2_{\rightarrow {\rm d}} \rangle \approx \frac{2}{\Gamma^2_{\rm sh}} - \theta^2_{\rightarrow {\rm u}}\,. \tag{22}$$

In both cases, it may be seen that the typical energy gain $\Delta E' \equiv E'_{\rm f} - E'_{\rm i}$ is thus of the order of $E'_{\rm i}$ itself [13].

Numerical Calculation of the Angular Distribution

As was seen in the case of non-relativistic shock acceleration, the power-law index of the accelerated particle distribution depends on the average energy gain per shock crossing and the escape probability. For relativistic shocks, both of these are strongly dependent on the angular distribution of particles crossing the shock, which as suggested in the previous section is in general highly anisotropic. Thus the quasi-isotropic approximations used in the non-relativistic case do not apply here, and the distribution of the shock crossing angles $\mu_{\rightarrow {\rm u}}$ and $\mu_{\rightarrow {\rm d}}$ has been evaluated numerically.

For simplicity, I will focus in this section on the case where both the upstream and downstream particle dynamics are dominated by scattering of the particle momentum direction. In other words, it is assumed that magnetic fluctuations dominate over the regular magnetic field in determining the particle transport, at least along the shock normal direction which is of interest here. This is a highly plausible assumption downstream, where shock-generated turbulence is likely to give rise to disordered magnetic fields significantly stronger than the shock-compressed upstream field, as is assumed in relativistic fireball models of gamma-ray bursts and their afterglows (see Sari & Galama, this volume).

One method of computing the accelerated particle distribution is through numerical simulations, and I illustrate below the results of such a calculation, after summarising the algorithm: since the nature of the particle transport upstream and downstream is by assumption independent of particle energy, it is computationally more efficient to decouple the dynamical problem from the energy gains. A numerical approximation to the function $f_{\rm d}(\mu'_{\rightarrow {\rm u}}; \mu'_{\rightarrow {\rm d}})$, the distribution of downstream egress angles $\mu'_{\rightarrow {\rm u}}$ for a given ingress angle, is thus first constructed by Monte-Carlo simulation of the downstream scattering process for a grid of $\mu'_{\rightarrow {\rm d}}$ values. The upstream dynamics are represented by a similarly obtained upstream egress angle distribution $f_{\rm u}(\mu_{\rightarrow {\rm d}}; \mu_{\rightarrow {\rm u}})$, and both distributions, along with the energy gain formula (20), are subsequently used in a Monte-Carlo calculation of the steady-state flux of accelerated particles crossing the shock.

The results of such a calculation were summarised in [14]. The influence of the highly anisotropic injected particle distribution was seen to disappear at

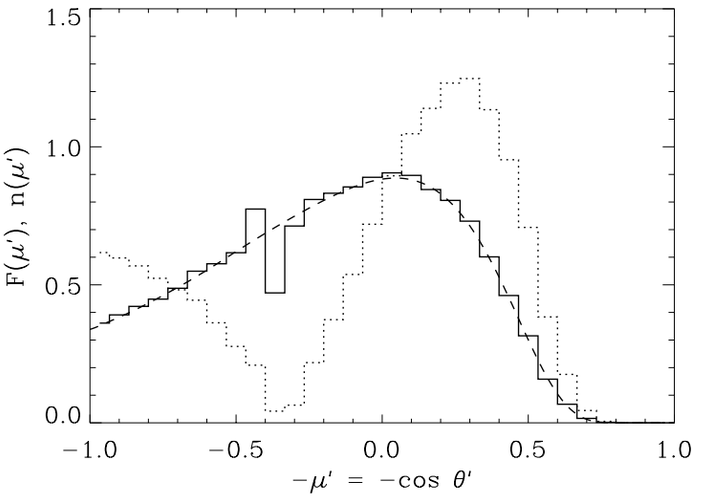

Fig. 3. Asymptotic downstream angular distribution of the particles crossing the shock, showing both the flux $F(\mu')$ (*dotted line*) and density $n(\mu')$ (*solid line*) obtained by Monte-Carlo simulations, along with the density obtained by the eigenfunction method (*dashed line*). All distributions are normalised to unity.

a little more than a decade above the downstream injection energy, at which point the self-consistent angular distribution was established with a power-law in energy, $F(E', \mu') \propto F(\mu')E'^{-p}$. Here F represents the steady-state flux of accelerated particles crossing the shock, per unit energy E' and direction angle cosine μ'. The asymptotic angular distribution obtained with this simulation method is displayed in Fig. 3, which shows both $F(\mu')$ and the corresponding density distribution, $n(\mu') \propto F(\mu')/(\mu' - \beta'_{\rm sh})$. The latter is compared with the distribution obtained with the very different semi-analytical eigenfunction method of Kirk et al. [21], showing excellent agreement between the two methods.

Spectral Index and Comparison with Observations

In the case considered above of isotropic scattering upstream and downstream of the shock due to a strongly turbulent magnetic field, a value of the spectral index $p = 2.23 \pm 0.01$ is found by both Monte-Carlo simulations [1,14] and the semi-analytical eigenfunction method [21]. For the opposite extreme in upstream dynamics of deflection by a regular magnetic field (see above), still assuming isotropic scattering downstream, Monte-Carlo simulations yield $p = 2.30$ [1]. With particle transport including both a regular field and magnetic fluctuations, Bednarz and Ostrowski [3] obtained in Monte-Carlo simulations values of $p \approx 2.2$ in the limit $\Gamma_{\rm sh} \gg 1$. A spectral index in the range $p = 2.2$–2.3 is thus a general feature of Fermi acceleration at (weakly magnetised) ultra-relativistic shocks, at least in the test-particle approximation used in all the above studies.

Spectral index values deduced from observations of astrophysical systems thought to involve ultra-relativistic shocks are consistent with these theoretical expectations. Early modelling of gamma-ray burst afterglow observations suggested $p = 2.3 \pm 0.1$ [32], and detailed analysis of the GRB 970508 afterglow spectrum yielded $p = 2.2$ [12]. While an equally detailed multi-wavelength spectral analysis has not been published for other afterglows, a value of $p \approx 2.2$ seems compatible with most [11]. In Crab-like supernova remnants, the inferred spectral indices are similar: the best-fit model for the Crab Nebula spectrum corresponds to p in the range 2.2–2.3 [19]. The very good agreement between theory and observation is all the more remarkable given that in non-relativistic shocks, as mentioned above, the observed particle spectra often differ from the predictions of the simple test-particle theory.

3.3 Moderately Relativistic Shocks

For moderately relativistic shocks, the self-consistent shock crossing angle distribution and the spectral index depend on the shock jump conditions assumed, as well as the shock Lorentz factor $\Gamma_{\rm sh}$, which together determine the shock velocity ratio. Figure 4 shows, for illustration, the spectral indices obtained with the eigenfunction method of Kirk et al. [21] in the two extreme cases considered in Sect. 2.3, namely that where the upstream gas is 'cold' so that the shock is strong, and that where it is ultra-relativistically hot.

It is readily seen that while for high Lorentz factors ($\Gamma_{\rm sh} \gtrsim 10$) the values obtained rapidly converge to the ultra-relativistic case, for lower shock Lorentz

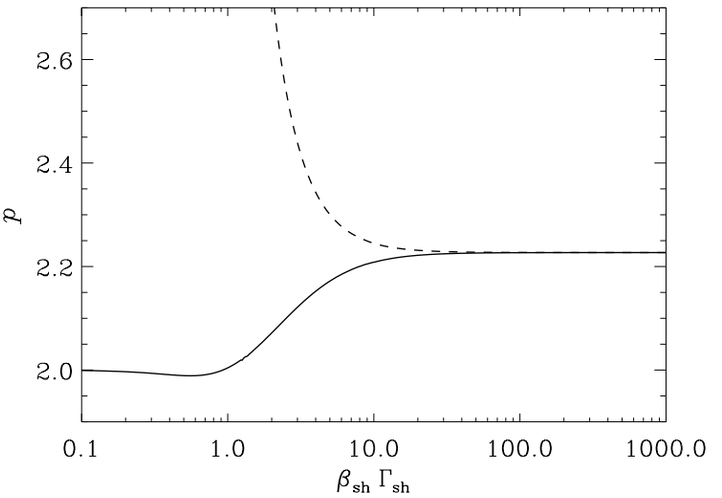

Fig. 4. Spectral index of the Fermi-accelerated particle distribution as a function of the shock four-velocity, for the two extreme cases shown in Fig. 1, namely a cold upstream gas (*solid line*) and an ultra-relativistically hot one (*dashed line*).

factors the different shock jump conditions yield very different spectral indices. In particular, for moderately relativistic shocks in a relativistically hot gas, which might be relevant to internal shocks in gamma-ray bursts, the spectral indices obtained can be significantly steeper than in the ultra-relativistic case. For strong shocks, on the other hand, the spectral index goes smoothly from the ultra-relativistic value of $p = 2.23$ to the non-relativistic one of $p = 2$ as the shock Lorentz factor decreases.

The above results apply to unmagnetised shocks and isotropic direction-angle diffusion both upstream and downstream. A non-negligible magnetisation parameter (4) lowers the strong shock velocity ratio, leading to steeper values of the spectral index than in the corresponding unmagnetised case [21]. Anisotropic diffusion in direction angle can have the opposite effect: calculations for the extreme case where the magnetic fluctuations are concentrated in the plane of the shock yield somewhat flatter spectra than the isotropic case, but by less than 0.1 in the spectral index p [21]. Anisotropic *pitch-angle* diffusion has been simulated by Bednarz and Ostrowski [3], who obtained steeper spectral indices for weak scattering, the regime considered above corresponding to the limit of strong scattering.

4 Maximum and Minimum Particle Energies

I now turn to the question of the range in particle energies over which the spectrum derived above applies. I first discuss the acceleration time scale for the Fermi mechanism, then use it to derive the maximum energy attainable by this process at a relativistic blast wave, and consider an alternative scenario involving the initial boost which can reach higher energies. I also discuss the minimum energy for Fermi acceleration, and the need for a distinct electron pre-acceleration mechanism in electron–ion shocks.

4.1 Acceleration Time Scale

The acceleration time scale $t_{\rm acc}$ is defined as the time needed for the particle energy to increase by an amount of order itself. Since, as seen in Sect. 3.2, this typically occurs every shock crossing cycle, $t_{\rm acc}$ is roughly the cycle time, which is the sum of the upstream and downstream *residence times* $t_{\rm up}$ and $t_{\rm dn}$. In the case of deflection by a uniform magnetic field upstream, the former is of order

$$t_{\rm up} \sim \frac{1}{\Gamma_{\rm sh}\omega_{\rm c\perp}} \equiv \frac{E}{q\Gamma_{\rm sh}B_{1\perp}c} \,, \tag{23}$$

where q and $\omega_{\rm c}$ are the particle's charge and cyclotron frequency. The downstream residence time depends on the downstream scattering process; assuming Bohm diffusion, it is roughly the downstream gyrotime,

$$t'_{\rm dn} \sim \frac{1}{\omega'_{\rm c}} \equiv \frac{E'}{qB'_2c} \,, \tag{24}$$

where primed quantities are measured in the downstream rest frame, as before. If the downstream magnetic field is simply the compressed value resulting from the (weakly magnetised) ultra-relativistic shock jump conditions, $B'_2 \approx B'_{2\perp} \approx \sqrt{8}\Gamma_{\rm sh} B_{1\perp}$, it can be shown that $t_{\rm dn} \sim t_{\rm up}$ [13].

Turbulence downstream may amplify the magnetic field B'_2 by a significant factor above the shock-compressed value, thereby reducing the downstream residence time by the same factor; assuming the field reaches equipartition with the thermal pressure downstream, this factor will be of order $c/v_{\rm A}$, where $v_{\rm A}$ is the upstream Alfvén speed. On the other hand, in the case of scattering by small-scale magnetic fluctuations, the upstream residence time is increased from the value (23) by a factor of order $R_{\rm L}/(\Gamma_{\rm sh}\ell)$, where ℓ, as before, is the correlation length of the magnetic field [13]. Thus the value of $t_{\rm up}$ given in (23) is a lower limit to $t_{\rm acc}$ in all cases of interest.

4.2 Relativistic Blast Waves and Ultra-high-energy Cosmic Rays

An immediate application of the above considerations is to the maximum energy attainable by Fermi acceleration at the relativistic blast waves occurring in fireball models of gamma-ray bursts; I thus first review some basic properties of these models. After an acceleration stage, an initially radiation-dominated fireball enters a relativistic 'free expansion' phase, in which a blast wave is driven into the surrounding medium with the approximately constant Lorentz factor $\Gamma_{\rm sh} \approx \sqrt{2}\eta$, where $\eta \equiv \mathcal{E}/(Mc^2)$, M being the baryonic mass in which the fireball energy $\mathcal{E}$ is initially deposited [24,28]. This is followed by an adiabatic deceleration phase in which the blast wave Lorentz factor decreases with radius $R_{\rm sh}$ as $\Gamma_{\rm sh} \propto R_{\rm sh}^{-3/2}$ [8]. The transition between these two phases occurs around the deceleration radius $R_{\rm dec}$, given by

$$R_{\rm dec} \approx \left(\frac{3}{4\pi} \frac{\mathcal{E}}{\eta^2 \varepsilon_1} \right) \ , \tag{25}$$

where $\varepsilon_1 \approx n_1 mc^2$ is the energy density of the surrounding material.

In the absence of energy loss processes, the maximum particle energy attainable by Fermi acceleration is set by the requirement that $t_{\rm acc}$ be shorter than the age of the system, which for a relativistic blast wave is simply $R_{\rm sh}/c$. Using (23) for $t_{\rm acc}$, the resulting maximum energy at a given blast wave radius $R_{\rm sh}$ is

$$E_{\rm max} \approx q B_1 \Gamma_{\rm sh} R_{\rm sh} \ . \tag{26}$$

Note that this is larger by a factor $\Gamma_{\rm sh}$ than a commonly used estimate resulting from a simple geometrical comparison of the particle Larmor radius with $R_{\rm sh}$ [17]. This is due to features specific to particle acceleration at a relativistic blast wave, in particular the fact that an accelerated particle typically executes only a fraction $\sim 1/\Gamma_{\rm sh}$ of a Larmor orbit upstream before recrossing the shock.

The evolution of the product $\Gamma_{\rm sh} R_{\rm sh}$ with $R_{\rm sh}$ implies that the highest $E_{\rm max}$ is reached at the deceleration radius, $R_{\rm sh} \approx R_{\rm dec}$. It has the numerical value

$$E_{\rm max} \approx 5 \times 10^{15}\ Z B_{-6} \left(\frac{\mathcal{E}_{52} \eta_3}{n_0} \right)^{1/3} \ {\rm eV} \ , \tag{27}$$

for particles of charge $q = Ze$, where B_{-6} is the upstream magnetic field, $\mathcal{E}_{52}$ the (isotropic) fireball energy, η_3 the initial Lorentz factor and n_0 the upstream density, respectively in units of microgauss, 10^{52} erg, 10^3 and cm^{-3}, these normalising values being those appropriate for a GRB fireball expanding into a generic interstellar medium [13].

Equation (27) rules out the production of ultra-high-energy cosmic rays (UHECRs), with energies up to $\sim 10^{20}$ eV, at the *unmodified, external* blast waves of relativistic fireballs. Scenarios which postulate a GRB origin for UHECRs [25,30,31] must thus invoke some other site or mechanism to reach the required particle energies. The idea most often put forward is that UHECRs are accelerated at internal shocks, where substantially higher magnetic fields, close to equipartition, could be present both upstream and downstream and allow the Fermi mechanism to reach UHECR energies. However, the particle spectrum in this case would likely be steeper, as argued above, reducing the efficiency of UHECR production. Moreover, the important issue of the escape of these particles from the interior of the fireball to the surrounding medium remains to be investigated, as they could suffer significant adiabatic losses due to the fireball expansion before escaping.

Another possibility for acceleration at the blast wave is that the upstream magnetic field might be amplified by a large factor above its undisturbed value due to instabilities driven by the accelerated particles themselves, as was recently proposed in the context of supernova remnants by Bell and Lucek [6]. It is unclear, however, how far such instabilities would have time to develop in the relativistic context, given the comparatively short time before the modified upstream medium is overtaken by the shock.

4.3 Fireballs in Pulsar Wind Bubbles

An alternative scenario for UHECR acceleration is based on the observation that the *initial boost* examined in Sect. 3.2 can circumvent the age limit (27), as it involves only a downstream half-cycle. The maximum energy is then set only by the requirement that the downstream residence time t_{dn} be less than the age of the system. If one assumes that the downstream magnetic field is turbulently amplified close to equipartition values, energies of order 10^{20} eV or more can be reached, provided that a population of relativistic particles with sufficient initial energy to be boosted into this range is present upstream [13].

Galactic cosmic rays with appropriate energies are present in the interstellar medium, but constitute only a small fraction of the upstream energy density, so that only a correspondingly small fraction of the fireball energy could go to boost these to UHECR energies. However, a situation where the surrounding medium consists almost exclusively of relativistic particles of the required energy occurs naturally in the context of neutron star binary merger events: the close binary pulsar systems observed in our Galaxy, which are the progenitors of these merger events, all contain millisecond pulsars with characteristic spindown times of order 10^8 yr, while their spiral-in times due to gravitational radiation are of order 3×10^7 yr [26]. These pulsars thus fill the surrounding space with relativistic

particles over the lifetime of the binary system, forming a large pulsar wind bubble in the interstellar medium.

While the majority constituents of these pulsar wind bubbles will likely be electron-positron pairs, pulsar winds also seem to contain ions [15,18]. Scaling this ion component to millisecond pulsar parameters, one can show that it yields ions with energies $\sim 10^{14}\,\mathrm{eV}$, sufficient to be boosted to UHECR energies provided $\Gamma_{\mathrm{sh}} \gtrsim 10^3$. This process is now highly efficient: a large fraction of the fireball energy can go to boost these ions to UHECR energies. Moreover, for typical parameters the blast wave will decelerate within the pulsar wind bubble, resulting in a power-law spectrum of boosted ions,

$$\frac{\mathrm{d}N}{\mathrm{d}E} \propto E^{-2} \,, \tag{28}$$

with a lower bound of $\sim 3 \times 10^{18}\,\mathrm{eV}$, compatible with the inferred UHECR source spectrum [13]. This scenario thus naturally provides for the acceleration of UHECRs into the required energy range, with the required spectrum, and with high efficiency.

4.4 Minimum Energy and Electron Pre-acceleration

Returning now to the Fermi acceleration mechanism proper, it has as one of its requirements that accelerated particles see the shock as a sharp discontinuity, as its treatment in Sect. 3 makes clear. For this to be the case, the particle Larmor radius must be larger than the shock thickness, which is in turn roughly given by the downstream thermal ion Larmor radius. Ions can thus undergo Fermi acceleration when they have reached a few times their downstream thermal energy, but electrons in electron–ion shocks must first reach a minimum energy

$$E'_{\mathrm{min}} \sim \Gamma_{\mathrm{rel}} m_{\mathrm{i}} c^2 \,, \tag{29}$$

where m_{i} is the ion mass, before participating in the Fermi mechanism and acquiring its characteristic spectral index.

Unless one assumes that these objects involve solely electron–positron shocks, the presence of synchrotron-emitting, Fermi-accelerated electrons in GRB afterglows and Crab-like supernova remnants thus requires an electron pre-acceleration mechanism. This mechanism must bring the electron energy from that resulting from randomisation of the bulk upstream energy, which as seen in Sect. 2.3 is $E'_{\mathrm{th}} = \Gamma_{\mathrm{rel}} m_{\mathrm{e}} c^2$, to E'_{min}, a factor of the mass ratio $m_{\mathrm{i}}/m_{\mathrm{e}}$ higher. An acceleration process operating over precisely this energy range is the resonant ion cyclotron wave absorption mechanism discovered by Hoshino et al. [18] in numerical simulations of highly relativistic, electron–positron–ion shocks.

This resonant ion cyclotron acceleration mechanism typically yields harder power-law spectra than those resulting from Fermi acceleration: the spectral indices p obtained from the simulations are generally less than 2, and a value as low as $p = 1$ is predicted in a quasi-linear, steady-state approximation [18]. The resulting picture for the accelerated electron (and positron) spectrum in

an ultra-relativistic shock containing ions is thus of a relatively hard power-law spectrum at low energies, steepening to the $p \approx 2.2$ spectrum characteristic of Fermi acceleration at a break energy given by (29). This might explain the flat radio spectral indices of Crab-like supernova remnants, as well as the two breaks in the Crab Nebula spectrum between radio and X-ray frequencies, only one of which can be attributed to synchrotron cooling.

5 Summary

The shock velocity ratio r across a relativistic shock is in general a function of the assumed upstream temperature as well as the shock Lorentz factor $\Gamma_{\rm sh}$, but it rapidly tends to the ultra-relativistic limit $r = 3$ for $\Gamma_{\rm sh} \gtrsim 10$. The ultra-relativistic Fermi acceleration regime then mirrors some of the simplicity of the non-relativistic, strong shock regime, this asymptotic shock velocity ratio corresponding to an asymptotic power-law index of the accelerated particle distribution. For the specific case of isotropic direction-angle scattering on both sides of the shock, this spectral index is $p = 2.23 \pm 0.01$; more generally, a value of p in the range 2.2–2.3 is found under a variety of particle transport assumptions. These values are consistent with the observed spectra of sources thought to contain ultra-relativistic shocks, such as gamma-ray burst afterglows and Crab-like supernova remnants. For moderately relativistic shocks, the spectral index depends on the shock jump conditions as well as $\Gamma_{\rm sh}$; in particular, shocks in a relativistic gas typically yield steeper spectral indices than the above ultra-relativistic values.

The maximum energy $E_{\rm max}$ of the Fermi-accelerated particle distribution is determined by the acceleration time, which is in general set by the upstream residence time. For acceleration at the unmodified, external blast wave of relativistic fireballs, this yields $E_{\rm max} \sim 10^{16}\,{\rm eV}$ for typical parameters of the surrounding interstellar medium, ruling out the production of ultra-high-energy cosmic rays in this context. If neutron star binary merger events give rise to relativistic blast waves with $\Gamma_{\rm sh} \gtrsim 10^3$, these can provide an alternative scenario for UHECR production: ions accelerated in the pulsar wind present before the merger can be boosted to energies $\gtrsim 10^{20}\,{\rm eV}$ by the blast wave with high efficiency; deceleration of the blast wave in the pulsar wind bubble yields a spectral index $p = 2$ and a typical lower cutoff around $3 \times 10^{18}\,{\rm eV}$. There is also a minimum energy for Fermi acceleration, set by the requirement that the shock thickness be small relative to the particle Larmor radius. In electron–ion shocks, this requires a distinct pre-acceleration mechanism for the electrons, which could be the resonant ion cyclotron wave acceleration mechanism of Hoshino et al. [18].

Acknowledgements

I gratefully acknowledge support from the Italian Ministry of University and Research through grant Cofin–99–02–02, and the Netherlands Organisation for Scientific Research (NWO) through GBE/MPR grant 614–21–008.

References

1. A. Achterberg, Y.A. Gallant, J.G. Kirk, A.W. Guthmann: M.N.R.A.S, submitted (2001)
2. W.I. Axford, E. Leer, G. Skadron: In: *15th International Cosmic Ray Conference, at Plovdiv, Bulgaria, August 13–26, 1977. Conference Papers*, Vol. 11, p. 132 (Bulgarian Academy of Sciences, Sofia 1978)
3. J. Bednarz, M. Ostrowski: Phys. Rev. Lett. **80**, 3911 (1998)
4. A.R. Bell: M.N.R.A.S. **182**, 147 (1978)
5. A.R. Bell: M.N.R.A.S. **182**, 443 (1978)
6. A.R. Bell, S.G. Lucek: M.N.R.A.S. **321**, 433 (2001)
7. E.G. Berezhko, D.C. Ellison: Ap. J. **526**, 385 (1999)
8. R.D. Blandford, C.F. McKee: Phys. Fluids **19**, 1130 (1976)
9. R.D. Blandford, J.P. Ostriker: Ap. J. **221**, L29 (1978)
10. E. Fermi: Phys. Rev. **75**, 1169 (1949)
11. D.A. Frail, T.J. Galama: personal communication (2000)
12. T.J. Galama, R.A.M.J. Wijers, M. Bremer, P.J. Groot, R.G. Strom, A.G. de Bruyn, C. Kouveliotou, C.R. Robinson, J. van Paradijs: Ap. J. **500**, L101 (1998)
13. Y.A. Gallant, A. Achterberg: M.N.R.A.S. **305**, L6 (1999)
14. Y.A. Gallant, A. Achterberg, J.G. Kirk, A.W. Guthmann: In: *5th Huntsville Gamma-Ray Burst Burst Symposium*, ed. by R.M. Kippen, R.S. Mallozzi, G.J. Fishman, p. 524 (AIP, New York 2000)
15. Y.A. Gallant, J. Arons: Ap. J. **435**, 230 (1994)
16. Y.A. Gallant, M. Hoshino, A.B. Langdon, J. Arons, C.E. Max: Ap. J. **391**, 73 (1992)
17. A.M. Hillas: Ann. Rev. Astron. Astrophys. **22**, 425 (1984)
18. M. Hoshino, J. Arons, Y.A. Gallant, A.B. Langdon: Ap. J. **390**, 454 (1992)
19. C.F. Kennel, F. Coroniti: Ap. J. **283**, 710 (1984)
20. J.G. Kirk, P. Duffy: J. Phys. G **25**, R163 (1999)
21. J.G. Kirk, A.W. Guthmann, Y.A. Gallant, A. Achterberg: Ap. J. **542**, 235 (2000)
22. J.G. Kirk, P. Schneider: Ap. J. **315**, 425 (1987)
23. G.F. Krymskii: Sov. Phys. Dokl. **22**, 327 (1977)
24. P. Mészáros, P. Laguna, M.J. Rees: Ap. J. **415**, 181 (1993)
25. M. Milgrom, V. Usov: Ap. J. **449**, L37 (1995)
26. R. Narayan, B. Paczyński, T. Piran: Ap. J. **395**, L83 (1992)
27. J.A. Peacock: M.N.R.A.S. **196**, 135 (1981)
28. T. Piran, A. Shemi, R. Narayan: M.N.R.A.S. **263**, 861 (1993)
29. J.L. Synge: *The Relativistic Gas* (North-Holland, Amsterdam 1957)
30. M. Vietri: Ap. J. **453**, 883 (1995)
31. E. Waxman: Phys. Rev. Lett. **75**, 386 (1995)
32. E. Waxman: Ap. J. **485**, L5 (1997)

Jet Formation and Collimation

Christophe Sauty[1], Kanaris Tsinganos[2], and Edoardo Trussoni[3]

[1] Université Paris 7 – Observatoire de Paris, D.A.E.C., F-92190 Meudon, France
[2] Department of Physics, University of Crete, GR-710 03 Heraklion, Crete, Greece.
[3] Osservatorio Astron. di Torino, Strada Osservatorio 20, I-10025 Pino Torinese, Italy

Abstract. We briefly review our current understanding for the formation, acceleration and collimation of winds to jets associated with compact astrophysical objects such as AGN and μQuasars.

All such outflows may be considered to a first approximation as ideal MHD plasmas escaping from a rotating and magnetized accretion disk with a magnetosphere around a central black hole. A crucial ingredient for a correct modelling of the steady state problem is to place the appropriate boundary conditions, by taking into account how information can propagate through the outflow and ensuring, e.g., that shocks produced via the interaction of the flow with the external medium do not affect the overall structure. As an example underlining the role of setting the correct boundary conditions, we make the analogy of the critical surfaces in the steady and axisymmetric MHD problem with the event horizon and ergosphere of a rotating black hole in relativity.

We discuss the acceleration of the outflow, by gas, radiation, or wave pressure gradients and also by magnetic mechanisms, showing the important role played by the disk corona in the vicinity of the black hole. Pressure and magnetic confinement both may also play a role in confining the outflow, although magnetic hoop stress confinement is likely to be a rather dominant process in tightly collimated outflows. The possible asymptotical morphology that jets achieve and the instabilities which are likely to explain the observed structures but do not prevent jets to possess toroidal magnetic fields are also reviewed.

Finally, it is proposed that in a space where the two main variables are the energy of the magnetic rotator and the angle between the line of sight and the ejection axis, some observed characteristics of AGN jets can be understood. A criterion for the transition of the morphologies of the outflows from highly collimated jets to uncollimated winds is given. It is based on the analysis of a particular class of exact solutions and may somehow generalize other earlier suggestions, such as the spinning of the black hole, the fueling of the central object, or the effects of the environment.

Thus, while the horizontal AGN classification from Type 0 to Types 1 and 2 may well be an orientation effect – i.e. a dependence on the viewing angle between the source axis and the observer as in the standard model – the vertical AGN classification with uncollimated outflows (radio-quiet sources) and collimated outflows (radio-loud sources) depends both on the efficiency of the magnetic rotator and the environment in which the outflows propagate.

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 41–70, 2002.

1 Introduction

1.1 Schematic Picture of AGN

Some galaxies are known to emit radiation with extremely high luminosities from a rather small volume in the $\gamma-$ray, X-ray and UV continuum. Such active cores are the so-called Active Galactic Nuclei (AGN) and the radiation is commonly believed to be a result of gravitational energy released by matter spiraling around a supermassive central black hole of about $10^9 M_\odot$ (see Fig. 1).

Though the central engine which produces the enormous observed activity cannot be resolved observationally, a standard picture of an AGN has gradually emerged to explain the richness of the radiation spectra:

- an accretion disk with radius from about 2 to 100 gravitational radii, R_g, feeding the central black hole and emitting mainly in the UV and soft X-rays;
- the broad line optically emitting clouds (BLR), which seem to be absent in some sources (e.g. FRI, see hereafter) and extend up to a few $10^3 R_g$ from the center. The BLR emission can be radiation scattered by hot electrons further away while the word "cloud" should be taken in the broad sense meaning dense gas with a filling factor less than unity [77];
- a dusty torus (or wrapped disk or dusty bipolar flow) with an inner radius of a few $10^3 R_g$, which obscures the central parts of the AGN from transverse lines of sight;
- the narrow line regions (NLR) which extend from about 10^4 to $10^6 R_g$;
- powerful jets of plasma detected from the sub-parsec to the Mpc scales, mainly visible in the radio but also in the optical, UV and X-rays.

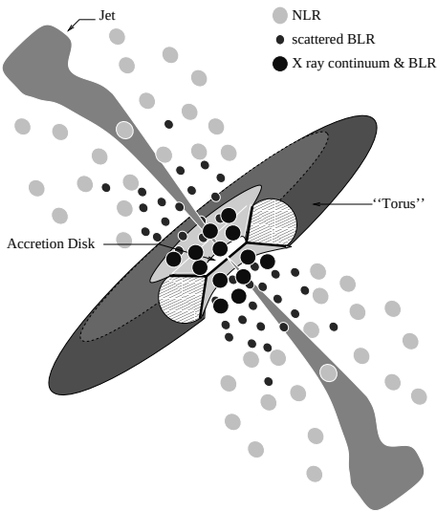

Fig. 1. General sketch, not to scale, of an AGN following Urry and Padovani ([98], see the text for details).

Note also, that ultra high energy γ-rays have also been observed from the central regions of several Blazars. Jets, together with the emission of radiation from the immediate neighborhood of an AGN provide a crucial link between the easier observed large Mpc scale and the sub-parsec scales where presumably the plasma of the jets is accelerated in a few gravitational radii from the center of the AGN.

1.2 Unified Schemes for AGN

Based on the phenomenology of their emission in the radio and optical/UV parts of the spectrum, AGN are commonly divided into three broad classes, as it may be found in excellent reviews in [1,98]. Although some details have been already modified since then, the overall classification still holds nowadays, at least as far as the properties of the associated winds and jets are concerned:

- Type 2 AGN have weak continua with narrow emission lines (NLR). They include, in the radio quiet group, the low luminosity Seyfert 2 galaxies and narrow emission line galaxies (NELG) while the radio-loud counterpart regroups the narrow-line radio galaxies with the two distinct morphologies of Fanaroff-Riley I (low-luminosities, FR I) and Fanaroff-Riley II (higher luminosities, FR II).
- Type 1 AGN have bright continua with broad emission lines (BLR) in addition to the NLR. The type 1 radio-quiet group is composed of the low luminosity Seyfert 1 galaxies (Sey 1) and the higher luminosity radio-quiet quasars (QSO), while the radio-loud group includes the broad line radio galaxies (BLRG) at low luminosities and the flat or steep spectrum radio-loud quasars (FSRQ and SSRQ) at higher luminosities.
- Type 0 AGN correspond to the remaining AGN with weak or unusual emission lines, i.e., in the radio-quiet end the broad absorption lines QSO (BALs) and in the radio-loud end the Blazars (BL Lacs and flat spectrum radio quasars, FSRQ).

Thus, the various AGN can be classified according to orientation, beaming and obscuration effects [98]. In this classification scheme, the transition from Type 0 to Type 1 and then to Type 2 of the class of radio-quiet AGN is based on orientation effects alone (Fig. 2). Namely, in Type 0 AGN the line of sight is almost coincident with the axis of the system, while the disk is seen face-on. In Type 2 the viewing angle is close to 90^o (disk edge-on) and obtains intermediate values for Types 1. The broad emission lines arise from "clouds" (i.e. dense gas with a small filling factor) orbiting above but nearby the disk (Fig. 1). Thus when the line of sight makes a small angle with the system axis, they are not obscured by the dusty torus, as in Seyfert 1, while wherein these broad emission lines are obscured by the torus, only the narrow emission lines are visible because they are produced further away, as it is the case with Seyfert 2.

Similarly among radio-loud galaxies, the transition from Type 0 (Blazars) to Type 2 (FR I/II radio galaxies) is based on a combination of orientation with relativistic beaming, i.e., whether a radio-loud AGN is a radio galaxy or a Blazar, and also depends on the angle between its relativistic jet and the line of sight. In this sense there seems to be a transition from FSRQ (Type 0) to SSRQ

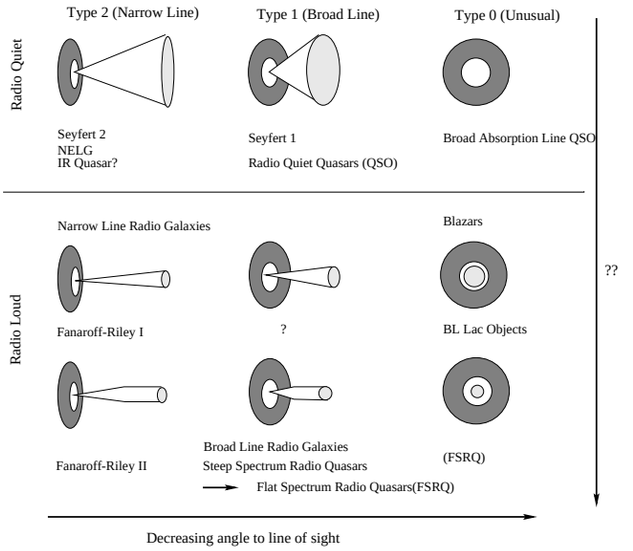

Fig. 2. Unified scheme presented by Urry and Padovani 1995 ([98]). Properties of AGN depend on at least two parameters: the viewing angle and some other parameter yet to be defined.

(Type 1) and then FR II (Type 2). For low luminosity radio loud galaxies there seems to be a gap as BL Lac objects (Type 0) are associated with FR I (Type 2) with no Type 1 counterpart. Although this association is still controversial, it may be explained by an intrinsic absence of broad emission line clouds [28] which would prevent to find any corresponding Type 1 objects with a BLR. This argument is supported by the fact that with increasing resolution BLR are also sometimes detected in FR II. At the same time however, recent data at optical and X-ray wavelengths have shown that, for the FR I/BL Lac case, the standard unification model does not seem to be in full agreement with observations. A possible way out to reconcile observations with the standard unification scheme is to assume a structure of the velocity across the jet ([21,29]).

It is now clear that orientation effects are not sufficient to explain the difference between radio-quiet, low luminosity and radio loud and high luminosity galaxies and quasars. It seems that in radio-loud AGN the outflow is relativistic at least in parsec scales, very well collimated in the form of a jet and quite powerful on large scales where it feeds the terminal radio lobes. Conversely, in radio-quiet AGN the outflow is either stopped or loosely collimated in the form of a wind or a bipolar flow. Parallely FR II jets are much more powerful than in FR I with a higher degree of collimation and terminal hot spots. Simultaneously the environment of the jets in FR I sources seems richer than in FR II ones. Various possibilities have been suggested. The radio-loudness could be related to: (i) the host galaxy type [90], (ii) the black hole's spin separating the lower spin radio-quiet galaxies from the higher black hole spin in radio-loud galaxies

[6,105], (iii) the differences in the rate of nuclear feeding [81,3], (iv) the different composition of the plasma [25], or (v) the different interaction with the ambient medium [43]. Nevertheless, none of these scenarios seem to be completely satisfactory because for all of them counter-examples may be found. We suggest at the end of this review a quantitative physical criterion for such a transition from radio-quiet to radio-loud galaxies which may, in fact, reconcile those various points of view by taking a different approach.

1.3 Towards a Similar Unification Scheme for μQuasars?

The galactic counterparts of the extra-galactic AGN were discovered recently by Mirabel and collaborators ([68] and references therein). Although the central black hole is not supermassive but just of the order of one solar mass, $M_{\odot}$, they also have relativistic ejecta with similar beaming effects. It is of course too early to draw a precise classification of such objects, since the number known so far is rather small in comparison to AGN. Nevertheless, there seem to exist μ-Seyferts and μ-quasars ([32,38] and Fender's review in this volume for details), with prototypes GX 339-4 and GRS 1915+105, respectively (Fig. 3). The winds of μ-Seyferts seem to be more continuous and conical while μ-quasars seem to have steady outflows in addition to pulsed collimated jets with higher speed ([35] and references therein). However, in X-rays these objects show also low/hard states where the ejection is present and a high/soft state where no outflow is produced, probably because of the disruption of the disk in the immediate vicinity of the central black hole. It is also interesting to note that especially in GX 339-4, the presence of the wind is associated with an extended X-ray corona at its base.

Note that we do not include in the present discussion all galactic relativistic jets from other binary systems but only those which have similar properties with AGN. Nevertheless, most of the mechanisms reviewed here apply also to such jets as they also apply, incidentally, to jets from young stars, stellar winds, etc.

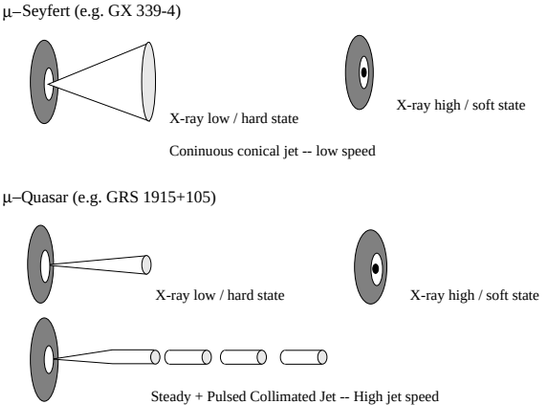

Fig. 3. Summary of the outflow properties of the galactic counterparts of AGN [38]. See text for details.

This may explain why the application of the theory of MHD winds, in jets from Young Stars and AGN has evolved parallely.

1.4 Some Key Problems about Jet Formation

The basic questions for understanding the physics and role of jets in AGN and μ-quasars and their complex taxonomy are those related to the nature of the constituting plasma, their initial acceleration in the environment nearby the central black hole, their morphology as they propagate away from the central region, the connection to the source (disk, disk corona or black hole magnetosphere) and extraction of angular momentum from it. Fig. 4 illustrates how the plasma is extracted from the magnetized rotating source, spiraling and carrying roped magnetic field lines.

Let us briefly recall here some points which we will not address in detail in the remaining part of this review.

First, it is likely that jets associated with quasars are a mixture of electron/proton and electron/positron pairs, since pure electron/positron jets overpredict soft X-ray radiation from quasars, while pure proton- electron jets predict too weak nonthermal X-ray radiation [89]. In fact, it has been suggested first by Sol et al. [88] that jets consist of electron/positron pairs close to their axis, surrounded by an electron/proton plasma (see Fig. 4). This model has shown some success since then [78,8] and it may naturally account for the presence of ultrarelativistic flows in the parsec scale made of pair plasma or a mixture of the two

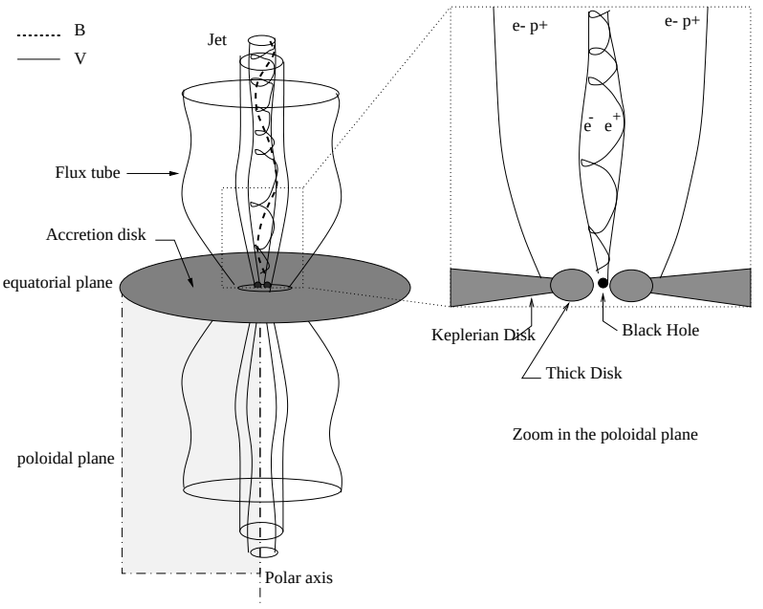

Fig. 4. View of the jet formation region.

while the lobes at larger scales would be fed with the proton-electron part. In this picture the lobes become illuminated by the faster inner jet at some point due, for instance, to an instability that disrupts the jet. It accounts also for the tremendous energetic power observed in the lobes which does not need to be the transported directly by the inner pair plasma.

Second, rotating outflows extract also angular momentum rather efficiently from their source, thus allowing mass to accrete on the central object. This removal of the angular momentum of the accreted material is very efficient in the presence of a magnetic field, as it is seen in stellar magnetic braking [86]. In fact, the magnetic lever arm ϖ_a at the point where corotation ceases is larger by a factor of order 10 than the cylindrical radius of the footpoint of a fieldline, ϖ_o (see Fig. 7b, [92,64]). It is interesting that even if a tiny percentage of order of 1% of the accreted mass rate $\dot{\mathcal{M}}_{acrr}$ is lost through a jet, $\dot{\mathcal{M}}_{jet}$, the major part of the angular momentum of the infalling gas is removed and so this gas can be freely accreted by the central object (see [92] for a simple explanation). However, though angular momentum extraction by the wind can be sufficient to account for the accretion (e.g. [40]), turbulent viscosity triggered by some instability can also allow for accretion. In this case, the outflow is likely to exist nonetheless (e.g. [23]). Thus the presence or absence of a jet is certainly not an evidence of what really triggers the accretion, though it certainly puts some strong constraint on it.

Let us just clearly state here that in the following sections we will mainly concentrate on the mechanisms which can produce *acceleration*, explain *collimation* and the related *nature* of the source in each case. We refer to various reviews [92,64,9,10] where complementary issues of the problem are treated in more detail. We are not addressing here the question of the propagation of the jet far from the source, its connection to the lobes and the external medium as this is the subject of other reviews in this volume (e.g. Aloy) and other contributions of the conference.

2 Basics of Jet Formation Theory

2.1 The Outflow MHD Equations

Since the outflows we are describing are made of tenuous electron/proton or electron/positron plasma they can usually be modeled to zeroth order via the set of the ideal MHD equations. Of course the collision rate may be so low that thermalization is not complete and each species should be treated separately. However, even in the well studied case of the solar wind where densities are probably even lower than in relativistic jets the fluid approach has proven to be very efficient and better than a pure collisionless one in describing the dynamics of the outflow. We shall not give here the general axisymmetric equations which can be found in the literature. Though it is an observed fact that at least some of the extragalactic outflows are relativistic in velocity or temperature, it has been shown that the basic physical mechanisms at work for the formation of jets are the same with those operating in the classical regime, e.g., [33,47]. Thus,

in the following we will not distinguish between relativistic and non relativistic approaches, unless some noticeable difference exists.

The full relativistic set of ideal MHD equations in the 3+1 formalism can be found in [19] for instance and their reduction to the classical Newtonian limit in [17]. They constitute a set of highly nonlinear and coupled partial differential equations of the four spatio-temporal variables. Note that in the following we shall use indifferently spherical (t, r,θ,φ) or cylindrical (t, ϖ, φ, z) coordinates. To describe the flow one needs then to determine its:

- mass density ρ,
- velocity field $\boldsymbol{V}$,
- magnetic field $\boldsymbol{B}$ (and electric field $\boldsymbol{E}$ in a relativistic treatment)
- gas pressure P (or equivalently, the temperature T) of the fluid.

This can be done by combining Maxwell's equations for the electromagnetic fields with the conservation of mass, momentum (Euler's equation) and energy for the hydrodynamic fields. The energy equation is usually (but not always) replaced by the simplifying assumption of a polytropic equation of state.

Under the assumption of steadiness ($\partial/\partial t = 0$) and axisymmetry ($\partial/\partial\varphi = 0$), the toroidal components (B_φ, V_φ) can be expressed in terms of the poloidal quantities [17]. Simultaneously, the magnetic field on the poloidal plane $[r, \theta] \equiv [\varpi, z]$ (Fig. 4) is defined by means of a scalar magnetic flux function A, $\boldsymbol{B}_p = (\boldsymbol{\nabla} A \times \hat{\varphi})/\varpi$ and the velocity field on the poloidal plane is defined by means of the mass flux function Ψ, $\boldsymbol{V}_p = (\boldsymbol{\nabla}\Psi \times \hat{\varphi})/\varpi$. Note that $\Psi = \Psi(A)$ because of the flux freezing law of ideal MHD. Practically, magnetic field lines and flow lines are roped on the same mass/magnetic flux tubes as shown in Fig. 4. Then the momentum equation splits in the poloidal plane into a component along each poloidal streamline and a component across it. Momentum balance along the poloidal flow (the Bernoulli equation) may be combined with momentum balance across the flow (the transfield or Grad-Shafranov equation) to form a system of two coupled partial differential equations for the density ρ and the magnetic flux function A.

Irrespectively of using a polytropic equation of state between pressure and density, or not, this system contains integrals that depend only on the magnetic flux distribution, such as:

- the mass to magnetic flux ratio, $\Psi_A(A) = d\Psi/dA$,
- the total angular momentum, $L(A)$,
- the angular velocity or rotational frequency of the footpoints of the magnetic fieldlines anchored in the wind source, star or disk, $\Omega(A)$, which is also the corotation frequency.

Note that $L/\Omega = \varpi_a$ must be the cylindrical radius of the field line A at the Alfvénic transition in order to ensure a smooth Alfvénic transition.

If a polytropic equation of state is used, an extra conserved quantity exists by integrating the momentum equation along the flow: this is the energy per unit mass, $E(A)$, which includes kinetic energy, enthalpy, gravity and Poynting flux. Usually the polytropic index is less than the adiabatic one (and even less than 3/2) in order to allow for thermal acceleration [76]. This is just a way to circumvent the solution of the rather difficult problem of solving the full

MHD equations by selfconsistently treating the heating supply in the plasma. In fact even if no polytropic assumption is made, a generalized form of the energy conservation can be written including a heating and cooling along the flow [84]. Some authors (e.g. [83]) prefer to use the energy in the corotating frame of rotation E' to show in this non Galilean frame explicitely the centrifugal potential. The two notations are equivalent and $E = E' + L\Omega$ where $L\Omega$ (called the "the energy of the magnetic rotator") is the energy a magnetic fieldline needs to corotate at frequency Ω and plays a crucial role in magnetic acceleration.

The remaining part of this section, despite that equations are not given, is rather more technical and the reader interested in the physical mechanisms at work may well skip it.

2.2 Axisymmetric and Time-Dependent Numerical Simulations

The time-dependent MHD problem has been treated only by means of numerical simulations for obvious technical reasons. Thus, there have been performed simulations of relativistic or non relativistic disk winds [75,49,100,101], outflows from a spherical magnetosphere [16,96,51,99] or, from both types of sources [52,55,70,56]. However, in some simulations it is not clear that the boundaries do not introduce spurious effects (e.g. [75,49]) or, that the system relaxes into a reproducible final state (e.g. [52,55,70,56]). Another difficulty with the numerical simulations has to do with the fact that AGN jets often extend over lengths more than six orders of magnitude their width, while the available grid sizes are much smaller. One way out of this constraint is to solve the problem via a combination of numerical techniques for the near zone and analytically solving the hyperbolic steady state problem at large distances from the central source (e.g. [16,96], although the very first accelerating region close to the base is not treated).

2.3 Axisymmetric and Steady Analytical Solutions

Several solutions of the steady MHD equations for various sets of boundary conditions are available analytically while there exists only one numerical solution for a specific and quite unique set of boundary conditions obtained by Sakurai [83] for stellar winds showing very weak collimation (i.e. logarithmically) around the rotational axis. Basically, the main difficulty is the fact that the set of the steady and axisymmetric MHD equations are of mixed elliptic/hyperbolic type, as opposed to the hyperbolic nature of time-dependent equations. Then, from the causality principle, a physically acceptable solution needs to cross three critical surfaces: the slow magnetosonic, the Alfvénic and the fast magnetosonic surfaces. However, the exact positioning of those critical surfaces is not known *a priori* but is only determined simultaneously with the solution. It is for this reason that only a few classes of such exact MHD solutions have been studied so far. They can be obtained by employing a separation of the variables (r, θ) in the poloidal plane (see Vlahakis and Tsinganos [103] for a general technique to obtain such solutions) combined with a suitable choice of the MHD integrals Ψ_A,

$L(A)$, $\Omega(A)$ and $E(A)$. It is worth to note that this systematic construction unifies all existing analytical models of cosmic outflows, such as the classical Parker wind [76] and the Blandford and Payne disk wind [18], in addition to uncovering new and interesting global models [103]. The best studied classes of such solutions are characterized by radial and meridional self-similar symmetries because all quantities scale with the spherical radius r or the colatitude θ respectively.

The *first* family with radially self-similar symmetry is appropriate to winds emerging from disks (e.g. [2,18,31,63,57,24,104,103] and references therein). A relativistic extension of these self-similar models exists (see e.g. [61]) although by dropping one essential element: gravity. No intrinsic scale length exists in this case and all quantities scale as a power law of the radius, similarly to the Keplerian law for the velocity in the disk. The key assumptions in this class of solutions are that the poloidal Alfvén Mach number M and the cylindrical radius ϖ of a particular poloidal fieldline A=*const.*, in units of the cylindrical radius ϖ_a at the Alfvén point along the same poloidal fieldline, are solely functions of the colatitude θ. In this case surfaces of constant M are assumed to be conical, and the critical surfaces too.

The *second* family is characterized by the meridional self-similar symmetry and is appropriate to winds emerging from a spherical source, although the physical variables are not spherically symmetric and the boundary conditions are functions also of the colatitude ([94,85,103] and references therein). Even though these solutions seem to be more natural to describe stellar winds, they do not exclude the presence of a surrounding accretion disk and they can on an equal footing describe a quasi spherical corona or magnetosphere around the central object. The key assumptions also in this case are that the poloidal Alfvén Mach number M and the cylindrical radius ϖ of a particular poloidal fieldline A=*const.* are solely functions of the spherical radius r. In this case surfaces of constant M are assumed to be spherical and so are the critical surfaces.

2.4 Boundary Conditions and Singularities

As we mentioned in the previous section, for the construction of a steady solution one has to carefully cross the appropriate critical surfaces encountered at the characteristic MHD speeds, corresponding to the three MHD waves propagating in the medium (but not to the elliptic/hyperbolic transitions as we discuss in the following). It effectively results in reducing the number of free boundary conditions [15]. Note that in the relativistic case the number of critical surfaces is the same with the nonrelativistic case because the light cylinder singularity is combined with the Alfvénic one [17]. As a corollary, this generalized Alfvénic singularity reduces to the classical one in the non relativistic regime and to the light cylinder when the mass loading is negligible.

It is conventional to use the crossing of the slow surface to fix the mass loss rate and the crossing of the Alfvén to fix the magnetic torque [92]. The crossing of the last singularity usually remains more subtle but is essential to ensure that no instability or shock with the external medium is propagating backward and possibly changing the structure of the solution. This is well known for the Parker

solar wind, where terminal shocks will naturally make breeze solutions to evolve into the wind solution.

To illustrate this point we show in Fig. 5 three disk wind solutions belonging to the radial self-similar class and the topology of another typical set of such solutions around the fast surface in the plane of colatitude and fast magnetosonic Mach number (instead of the Mach number for a classical hydrodynamic wind). A careful look at the behaviour of those solutions shows that one is crossing exactly the fast magnetosonic transition [104] while the other two behave close to the critical surface either like a breeze solution [40] or like a terminated solution [24] in Parker's terminology for the solar wind, as illustrated on Fig. 5. Although they differ in the way they connect to the underlying disk, they basically show very similar properties. This indicates that, by tuning the heating deposition along the flow or the polytropic index, a physical connection between the disk and the crossing of the fast point is possible despite it has not been done yet. It also suggests that the crossing of the last surface is not so crucial for the connection with the disk but it validates, nonetheless, the widespread use of such disk wind solutions. In fact all solutions terminate after some point due to the presence of a spiral singularity as shown in Fig. 5, which makes even more crucial the presence of a shock not only from observational arguments but also from theoretical ones.

2.5 Singularities and Horizons

We discuss now the true nature of the singularities, pointing out that a strict analogy between MHD signal propagation and the light propagation in the neighborhood of black holes (see Carter [22]) exists.

First, as it is well known in the theory of steady and spherically symmetric black holes (Schwarschild black holes in the case of vacuum), the horizon where no information or light signal can escape coincides with the ergosphere where

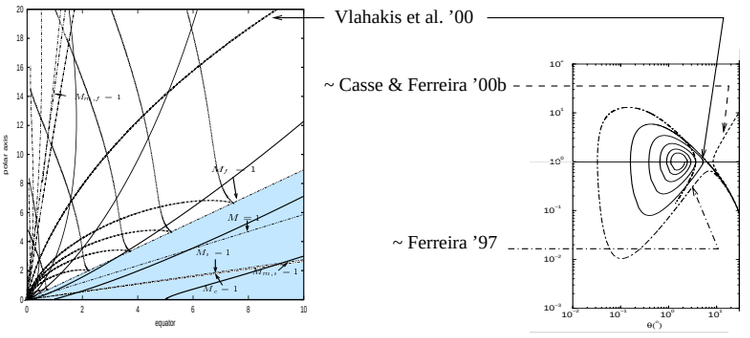

Fig. 5. Topology of an analytical disk wind solution in the region of the fast magnetosonic transition in the $[M_f,\theta]$ plane. Examples of three solutions of this type are drawn: one that crosses the critical point and two which do not.

the system of the relativistic equations changes nature, from elliptic outside the horizon to hyperbolic inside it. As illustrated in Fig. 6a the system is hyperbolic in time and information emitted at the speed of light propagates along a "cone", the light-cone. In a steady state and far from the black hole, the light-cone's projection is a circle and thus information can propagate in all directions of the X_1 - X_2 plane, similarly to water waves on a static pool, where X_1 and X_2 are two coordinates in space. The equations are elliptic. But because of gravity, light is deflected and inside the horizon, equations change to hyperbolic, the cone projection gives characteristics, as the trail of a boat, and information can propagate only inside these. Furthermore, all characteristics converge to the center of the black hole such that no information can escape (Fig. 6a).

The same is true and well known in MHD outflows where light is replaced by MHD waves and the spatial coordinates are reversed (X_1 corresponding to $1/X_1$) such that the center of the black hole becomes the asymptotic outer connection of the MHD outflow with the extragalactic medium. In spherically symmetric outflows (or equivalently when the poloidal geometry of the flow is fixed), each of the three singularities coincides in fact with a spherical surface which marks the transition in the nature of the equations from hyperbolic to elliptic[1].

[1] There is an extra transition at the cusp velocity which is not a singularity while the Alfvén singularity is in a parabolic domain but we shall not enter here in these details.

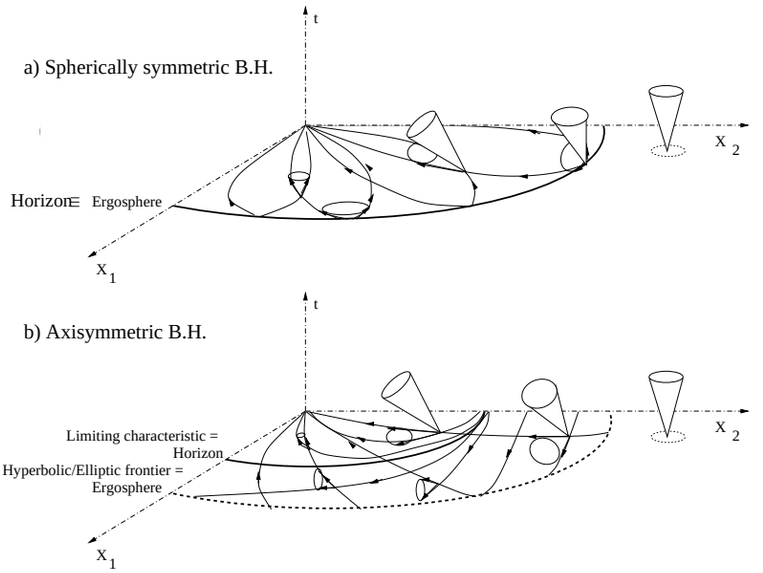

Fig. 6. Ergosphere, horizon and light cones around a black hole after Carter [22]. Time is along the vertical axis and space in the horizontal plane $[X1, X2]$. In a) sketch of a spherically symmetric black hole and in b) a rotating axisymmetric one.

Second, if the black hole is rotating, the event horizon and the ergosphere where the system changes from elliptic to hyperbolic split. It is easy to understand this physically (see Fig. 6b). As the geodesics rotate, the light cone first inclines itself such that beyond the ergosphere characteristics appear. One of the family of characteristics still connects with the external medium so it is still possible to propagate information backward. Once inside the real horizon the light cone is inclined and directed towards the center. At this point the outer space is causally disconnected from the interior of the horizon.

The analogy between the event horizon and the MHD singularities in terms of limiting characteristics or separatrices has been recently recognized [15] and illustrated explicitly by examples of self-similar solutions [95]. Note that for the slow magnetosonic horizon the situation is a bit more complicated because the slow waves have triangular wave fronts instead of circular [102]. However, the ultimate horizon is the one associated with the fast waves. In addition, the so-called "classical" critical points correspond to the ergosphere where the equations change from elliptic to hyperbolic. This point has been underestimated. In fact, it is known for the ergosphere of black holes that it can appear as a singularity but a suitable choice of the Killing vectors eliminates it. The same must be true for the "classical" critical points where the poloidal velocity equals one of the wave speeds.

Third, the analogy can still be pushed one step further. There are only two cases where the horizons can be defined locally, either, for static black holes where the horizon coincides with the ergosphere (Fig. 6a), which corresponds to the spherical Schwarzschild solution in vacuum, or, in the circularity limit where the horizon coincides with the rotosurface, which means that there is only rotation and no convection. In all the other cases, the horizon can be defined only globally that is as a limit of the characteristics, once the solution of the metrics is known (i.e. as a limiting characteristic).

The circularity limit for the vacuum solution corresponds to the Kerr rotating black hole (Fig. 6b). In this special case, one can construct the solution [22], by means of a separation of the variables r and θ while t and φ are ignorable. The θ component can be solved by using Legendre's polynomials. This way of constructing the global solution is identical to the one used for self-similar flows. In both cases the form of the singular surfaces is known a priori and this is definitely NOT the self-similar assumption that "modify" the critical points as we have written for so long.

The first conclusion is that in the more general axisymmetric case, there is little hope that we can determine the limiting characteristics *a priori* in MHD outflows. However, these are the true singularities of the flow.

The second conclusion is that this problem of horizons concerns not only steady solutions but also numerical time-dependent solutions. Even if they do not appear as real mathematical singularities, they must be present in the sense that characteristics on the outer boundaries should all be directed outwards, which has not been the case of all simulations as we already mentionned.

3 Acceleration

Once valid solutions of the outflow equations are obtained, either numerically or analytically, we may study the physical mechanisms that accelerate and collimate the outflow transforming it from a wind to a jet. From a rather general perspective, an observational characteristic of many cosmic plasma outflows is that they seem to be *accelerated* to relatively high speeds which may even reach values close to the speed of light in the most powerful AGN jets. The most often invoked mechanisms to accelerate these flows are of thermal, or, of magnetocentrifugal origin. We shall discuss first in the following magnetocentrifugal acceleration since it is widely considered as the most relevant mechanism for the acceleration of AGN jets.

3.1 Toroidal Magnetic Field Acceleration

The simplest magnetic driving mechanism is the so-called 'uncoiling spring' model (Uchida and Shibata 1985, [97,56]) or 'plasma gun' [30] where a toroidal magnetic pressure builds up due to the rotation of the fieldlines which are anchored in the disk. Evidently, there is a net force pushing the plasma upwards, as shown in Fig. 7a. This mechanism is mainly seen in numerical simulations (e.g., [87]). After the initial transient phase when a torsional Alfvén wave develops and drives the initial acceleration of the flow, the solutions converge to a weakly collimated structure, where the confinement is done by the toroidal magnetic field. However, for the numerical constraints we mentioned above, the outflow cannot be simulated in regions far from the base to follow realistically its degree of collimation. Second, such numerical simulations were able to follow the jet for one or two rotations around the central body. Hence, such a mechanism seems to be at work only to explain intermittent ejection, something equivalent to Coronal Mass Ejections (CMEs) in the solar wind that travel on top of a global more steady structure.

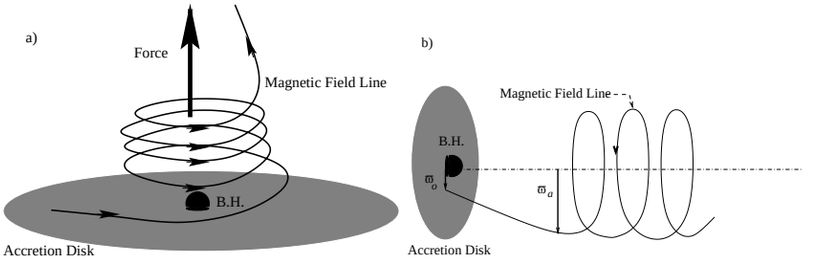

Fig. 7. a) Acceleration by toroidal (azimuthal) magnetic pressure. The fieldline is wounded by rotation and acts as an uncoiling spring. **b)** Magnetocentrifugally driven wind. The acceleration is similar to that of a 'bead on a wire' and it operates from the disk footpoint ϖ_o up to the magnetic lever arm ϖ_a where corotation stops. Further downstream the magnetic field is rapidly wound up and magnetic collimation is obtained because of the pinching magnetic tension (after Spruit [92]).

It is interesting that instabilities by a spiral wave in the disk producing an Alfvén wave have been also advocated to explain the intermittent ejection from μQuasars[93]. These instabilities can be at work to explain various features, as we briefly shall discuss later and transients on a time scale of a few rotations around the black hole. However, we need to model at the same time some more continuous ejection, similarly to the solar wind where despite instabilities and CMEs, there always exists a steady wind outflow.

3.2 Magneto-centrifugal Acceleration from the Disk

In order to accelerate jets magnetically, the most popular scenario is the magnetocentrifugal acceleration from an accretion disk via the classical Blandford & Payne [18] acceleration mechanism. In this case, the plasma consists probably mostly of electrons/protons, either relativistic or not. Electrons/positrons could be accelerated in the same way but it seems more difficult to produce them above the keplerian disk than in or close to the black hole's magnetosphere for energetic reasons, e.g. [46]. As the poloidal magnetic field is dominant up to the Alfvén radius, it practically enforces an approximate plasma corotation similarly to a 'bead on a rotating wire' (see Fig. 7b). In the corotating frame there is then a centrifugal potential which accelerates the flow outwards provided that the line is sufficiently inclined ($\theta > 30^o$ from the pole for a cold non relativistic plasma). Note that this condition is less restrictive if there is some heating or the gas is relativistic [91,92]. This point should not confuse the reader, it is the combination of centrifugal force and a strong poloidal $\boldsymbol{B}$ field that allows for the acceleration to take place. Eventually the acceleration comes from the conversion of Poynting energy flux to kinetic energy flux. The gain in kinetic energy is proportional to the energy that brings the magnetic field lines into rotation, i.e., the energy of the magnetic rotator $L\Omega$. Such a magnetocentrifugal driving mechanism seems to be efficient in disk winds wherein a hot corona is not an absolute requirement.

Besides a long list of self-similar models (see [103,53,54,74]), the same mechanism has shown to be successful in various numerical simulations [75,49].

This mechanism has some limitations. For example, it requires high magnetic field strengths at the disk level (which have not been measured so far) and also a large magnetic lever arm is needed in order to obtain a terminal speed which is a few times the Keplerian speed ($\lesssim 10^5$ km/s). Moreover, only a very small fraction of the accreted mass can be ejected at very high speeds, once a connection with a realistic disk structure is properly made [40,63,23]. By realistic disk structure we mean that in the accretion disk resistivity, turbulence and viscosity are taken into account within the hypothesis that the disk is pervaded by a large scale mean magnetic field. However, the presence of a hot corona on top of the disk is enough to eliminate this limitation [24]. If these conclusions hold unchanged in the presence of a local dynamo and/or disordered magnetic field is not clear yet [45,10].

3.3 Magneto-centrifugal Acceleration from the Black Hole Magnetosphere

A wind outflow could also be extracted "magneto-centrifugally" from a black hole's magnetosphere through the Blandford & Znajek mechanism [13] (see also [10] for more references). First, the rotational energy of the accreting black hole is extracted by a large scale magnetic field accreted onto the black hole, then converted to Poynting flux and finally to relativistic electron/positron pairs. As far as the plasma itself is concerned there is no difference with the previous mechanism, since it is ultimately the Poynting flux that accelerates it. However, there are two basic differences. First, the Poynting flux is extracted from the black hole. This is physically consistent since some angular momentum of the opposite sign is lost simultaneously into the black hole. Second, there is enough energy in the magnetosphere to produce electrons/positron pairs, such that this mechanism favors leptonic ejection. It has been argued that pairs would suffer Compton drag; however, this may be a real problem for radiatively driven winds but not if there is an extra mechanism to accelerate the flow sufficiently to overcome these radiative losses.

The efficiency of this mechanism has been recently put into question by several authors (see for instance [65]). The main argument is that the extraction of energy from the black hole through this process is at most as efficient as the extraction of energy from the disk. This only means that the two mechanisms are likely to operate simultaneously. Again this could very well be in favour of a leptonic jet or beam extracted from the back hole embedded in a hadronic heavy wind/jet coming from the disk [88,46]. Moreover this does not apply necessarily to the gamma ray emission which may still get its energy from the black hole [10].

3.4 Radiative Acceleration

The first alternative to a magnetically driven wind is a radiatively driven wind. For instance, in the 'Compton rocket' model a disk produces electron/positron pairs of which some are accelerated by the radiation produced by the annihilation of this plasma. Although Compton radiative losses exist, they are not sufficient to prevent completely the plasma acceleration [71,27].

Radiatively driven models for disk winds of electrons/protons have been proposed with a radiative pressure due to dust or, due to line emission coming either from the disk or the central source (see [79] and further references in the introduction). The key point in such models is that they result in at most a few tens of thousands of km/s for the outflow speed ($\sim 50,000$ km/s). This may be enough to explain most of the winds from radio quiet AGN, but it is unable to explain the acceleration of the powerful jets associated with radio loud AGN, or, the mildly relativistic flows seen in some radio quiet AGN. It is thus likely that in those objects, radiative acceleration may operate as a minor contribution or, in combination with other mechanisms like magnetocentrifugal driving [53].

Alternatively, it has been shown that a magnetized cloud of relativistic electrons can be radiatively driven up to relativistic bulk velocities if the interaction between photons and particles comes through a synchrotron process [41,42]. However the cross section of this mechanism is critically dependent on the geometry of the interaction and some assumptions are necessary 'a priori' for the treatment of the equations.

3.5 "Thermal" Acceleration

The other alternative to magnetic acceleration is classical thermal driving, as it is the case in the low- and high-speed solar wind where the heating and part of the pressure is provided by the dissipation of acoustic waves, electric currents, etc., or more efficiently by Alfvén waves. In this case the presence of a hot corona around the disk and/or the magnetosphere is essential for the acceleration, which is proportional to the sound speed, i.e., to the square root of the coronal temperature.

Then, in a rather crude estimate, if the 10^6 K corona produces a thermally driven wind with a terminal speed around 300 km/s, a corona with a temperature of 10^9 K for both the ions and the electrons could result in a terminal speed around 10,000 km/s. On the other hand, if 10^9 K is the temperature of the electrons while the temperature of the protons is 10^{12} K, a wind results with a terminal speed of the order of the speed of light, 300,000 km/s. Of course at this point relativistic effects should be taken into account properly. For ultra-relativistic flows the adiabatic sound speed is only $c/\sqrt{3}$; however, this is without taking into account the existence of extended heating in the corona. Most of all the heating by waves, in particular (torsional) Alfvén waves, could be very efficient. For instance, it has been proven to be efficient enough, even with small amplitudes, to explain the 800 km/s of the fast solar wind [99]. By extrapolation, we may guess that it should be able to produce very high speeds in AGN outflows. Waves of large amplitudes could be even more efficient, producing turbulence, and this is not very different from the transient torsional Alfvén "wave" seen in numerical simulations [56], except that the production of such waves should be continuous.

In fact, this mechanism combined with magnetocentrifugal driving, has found some success in the literature [64], like for the acceleration in the corona from ADIOS [7], from a Keplerian disk [24], from a black hole magnetosphere through a shock [34,52,55] or, more in general, from any kind of spherical corona [84,94,85].

Thus, thermal acceleration in a broad sense is likely to be as efficient as the magnetic processes. We could then suggest that both may be at work in disk winds for electron/proton plasmas (where it also allows to have higher mass loss rates, as we already mentioned [24]) while the electron/positron pairs would be more likely magnetically driven from a black hole magnetosphere. This seems also to be suggested by recent numerical simulations [52,55] in which two flow streams are accelerated near the black hole after passing through some shock in the accretion disk. Despite the fact that the disk is governed by ideal MHD and it is not clear if the simulations can hold more than one or two rotation

times, it is worth to note that these simulations show a double component of the wind, a inner magnetocentrifugally driven part and an external pressure driven component.

4 Collimation

Once the outflows are accelerated, they will propagate in the form of either collimated beams or uncollimated winds. However, apart for the case of the solar wind, uncollimated flows are difficult to observe, while jets can be seen in several astrophysical environments, from star formation regions to distant AGN. This is mainly due to the much higher density inside jets as opposed to that in loosely collimated winds. Furthermore, in radio-loud AGN where beams move at relativistic speeds, the emission may be largely amplified by Doppler boosting if the jet is pointed towards the observer, which is probably not the case in radio-quiet sources. Again the two basic mechanisms responsible for collimation may be of thermal or magnetic origin.

4.1 Pressure Confinement

An outflow is thermally confined if the surrounding medium has a higher pressure than the flow, such that there is a pressure gradient forcing the outflow to collimate along its ejection axis. In other words, only outflows underpressured with respect to their surrounding environment may be thermally confined. In fact, such a situation seems to occur in many extragalactic jets, as deduced from X-ray data implying a hot plasma surrounding early-type galaxies and clusters of galaxies [39].

The 'twin exhaust model" based on an analogy with the De Laval Nozzle, was the first effort to thermally confine jets [12]. However, this confining mechanism has been by now excluded because it requires the throat of the nozzle to be located rather far from the central object, as it works for both collimating and accelerating the flow. To remedy that, the idea of an external medium only collimating the outflow has been suggested [36]. Meridionally self similar models [94,85] have shown that cylindrical collimation could arise naturally from inward pressure forces but with some contribution by the magnetic field, too. In a pure thermally collimated flow the jet finally should collapse onto the rotational axis, unless it rotates fast enough such that the centrifugal force may counteract the external pressure. However, the strength of rotation is likely to be related to the strength of the magnetic field (because of dynamo effects for instance) such that a low magnetic field would indicate a weak rotation.

So thermal confinement may play a role in FR I and Seyfert types of AGN, but probably not in FR II, which are known to have a very poor environment, so there cannot definitely be a unique mechanism for all AGN.

4.2 Poloidal Magnetic Confinement and Subfast Flows

With toroidal magnetic fields known to be unstable in tokamaks, it has been suggested that toroidal confinement and pinching should be unstable, and collimation could be achieved by poloidal magnetic fields alone [91]. This is supported by the parallel magnetic field measured at small parsec scales in some extragalactic jets [92]. However recent observations of highly optically-polarized compact radio-loud quasars (HPQ) have shown that the electric vectors of the polarized 43 GHz radio cores are roughly aligned with the inner jet direction indicating magnetic fields perpendicular to the flow [66]. On larger scales, magnetic fields are also known to be perpendicular to the jet axis in FR II sources while they are parallel to the jet axis in many FR I sources. However parallel does not necessarily mean that it is not helicoidal and there is no toroidal field. The parallelism could simply be due to a strong velocity shear across the jet's cross section as it is explained in [10].

Poloidal magnetic fields can induce some mild collimation in the outflow in the region from the disk up to the Alfvén transition (see Fig. 7b) where the plasma in the transfield equation governing the morphology of the flow is basically dominated by magnetic forces. Beyond this distance, the jet becomes superalfvénic and the hydrodynamics of the flow overcome magnetic forces, in such a way that collimation will stop. To continue poloidal collimation on large distances, the jet must remain subalfvénic. But then, the jet will be very sensitive to shocks and instabilities that can propagate upstream from far distances and destroy the whole equilibrium.

Similar problems occur in asymptotically cylindrical solutions [31,74] of radially self-similar disk-wind models. Subfast outflows as those proposed by Ostriker [74] attain only low Alfvén Mach numbers and such solutions are structurally unstable [104]. In fact, as we already wrote, all the other solutions of those models are terminated because of the spiral singularity (see Sec. 2.4).

4.3 Toroidal Magnetic Confinement and Stability

Another confining mechanism, which is in fact supported by observations [66], is the magnetic confinement of the outflow by a toroidal magnetic field wound around the jet, the so-called hoop-stress paradigm [92]. This mechanism works both for under- and over-pressured jets. Observations of perpendicular magnetic fields [66] imply that such beams carry some electric current that eventually closes at their surface or outside. Note that the building of the toroidal magnetic field is done at the expense of the Poynting flux. Thus, in a pure magnetic jet all the Poynting flux cannot be converted to kinetic energy, if part of it remains to confine the jet. Obviously reality in most cases, and particularly in the case of AGN, may involve a combination of thermal and magnetic processes in the acceleration and confinement of the outflow.

Despite the observations that we mention above, there is still a vigorous debate on whether magnetic instabilities may ultimately disrupt the jet, or not. In

particular, it has been shown in the context of a pure magnetic jet without rotation that instabilities are always present [4]. Rotation in a pure hydrodynamic flow is also known to have a destabilizing nature and hydrodynamic instabilities may also disrupt the jet [14], though relativistic jets are more stable (see Aloy this volume). However, the combination of toroidal magnetic fields and rotation is more subtle, since the two ingredients act in opposite directions.

Recent numerical simulations and extended analytical work [48] gave support to the result that magneto-rotational instabilities may develop rapidly. Such instabilities for a cold plasma tend to favor the formation of an inner denser core in a jet. Similar results have been obtained from a local analysis of the ballooning modes [50], though there is no precise calculation of any growth rate. In particular, the inner part of the jet with a vanishing current density on the polar axis, is particularly unstable to magnetic shearing in this analysis.

On the other hand, a non linear analysis of current driven instabilities ([59] and ref. therein) has shown that the instability instead of disrupting the jet leads to a reorganisation of the current density. Instabilities in more complex jets from Keplerian disks have also been studied [60] and present interesting structures, again forming a dense core jet surrounded by a return current in a cocoon.

Parallely, magnetorotational instabilities have also been studied using the flux tube approximation [44], which is likely to give the most unstable mode, including only toroidal fields but with all the other ingredients, such as shearing or buoyancy. This study agrees with the fact that jets with a Keplerian velocity profile, as well as the outer region where the jet connects to the external medium, are subject to strong instabilities. Conversely, the inner parts of a jet can be completely stabilized for a flat, or a solid rotation profile, provided – and this does not appear in the studies mentioned above – that there is an increase of the density away from the axis. In this case jets from the central source, like those obtained in meridional self similar models, should be more stable than jets from Keplerian disks. It also explains the edge brightening seen in some sources, because the instabilities are likely to occur mostly at the edges. Eventually instabilities are sources of reacceleration in the plasma and ultimately radiation. Finally, it suggests that hollow jets (not empty jets!) should be more stable than dense core ones.

Altogether, then it follows from both, observations and theoretical studies, that one better be careful before claiming that toroidal instabilities will disrupt rotating MHD jets. However, instabilities do in general exist and they are obviously a crucial element for explaining a few structures we see in both observed jets and simulations. The issue is rather crucial and is a subject by itself in this volume (see the article by Aloy).

4.4 Asymptotic Equilibria

With hoop stress and pressure gradients collimating winds into jets, it is interesting to wonder which kind of asymptotics the outflow takes. A useful general analysis for magnetically dominated flows has been performed by Heyvaerts & Norman [47], generalized to relativistic flows in [33].

In the case where pressure and centrifugal forces drop asymptotically faster than the magnetic forces do, then the final equilibrium state should be force-free, and the asymptotical morphology of the flow is related to the electric current flowing. To summarize, a given flux tube collimates to

- *cylindrical* asymptotics if there is a net poloidal current spread in it;
- *paraboloidal* asymptotics if the net poloidal current is zero;
- *radial* asymptotics if there is a net current but it flows inside the region of the radial asymptots.

Of course, the whole jet could be in principle cylindrical if the return current lies outside the jet or, in a current sheet (cf. observations [66]). Note that flows with return current sheets are known to exist and the best example where it is observed and measured *in situ* is the solar wind. However from the analysis of Heyvaerts & Norman (1989) the possibility of mixed asymptotics with cylinders surrounded by cones (=radial asymptotics) is not excluded. Asymptotic solutions of such flows have been successfully constructed especially in the context of relativistic flows either for pure cylindrical asymptots [37]) or mixed radial and cylindrical (e.g. [69],[16]).

An analogous analysis to [47] has been recently proposed, which nevertheless arrives to opposite conclusions [72]. It extends the study to the quasi-asymptotic domain (called asymptotics), i.e. for $z \lesssim \infty$ and connects the curvature of the streamlines to the direction of the current density. It is shown that the collimating part of the outflow is related to the enclosed current, while the outside region of return current should cause the outflow to decollimate. However, it is postulated that cylindrical and radial asymptotics are not accepted as a valid possibility in MHD outflows because of their "violation of causality" and because cylinders correspond to a specific direction, as claimed therein. Instead, a continuous deflection towards the polar axis or the equatorial plane is preferred. However, this inevitably leads to an inconsistently infinite density there while close to the polar axis the underlying assumptions are not valid any longer as we explain below. A more interesting relativistic generalisation of these results is obtained in [5] where collimation is not rejected as a possibility but it is pointed out that the presence of decollimated flows could explain strong equatorial flows seen around several compact objects.

All previous results rely on the strong hypotheses that pressure and centrifugal forces drop asymptotically. In order to have this, two assumptions are made. First, that the cylindrical radius of any flux tube is assumed to be much larger than its value at the Alfvénic transition $\varpi/\varpi_a \to \infty$ – this is not necessarily true for cylindrical asymptots even for $\varpi \gg \varpi_a$. Second, it is assumed that we do not remain too close to the polar axis [72]. More general asymptotics have been found [84,94,85] including pressure and rotation but using the assumption of meridional self-similarity which cannot hold indefinitely far from the axis (where the return current is well known to exist for instance). Cylindrical and radial asymptots are found in agreement with the Heyvaerts & Norman [47] conclusions despite the different assumptions.

Then, after considering the asymptotic behaviour in general, one needs to connect it to the source [37,58] and if possible by solving selfconsistently the transfield equation [84,94,103,85,16,96], a topic we take up in the next section.

5 On a Possible Classification of AGN

5.1 An Energetic Criterion for the Collimation of Outflows

Apparently a missing parameter in the vertical classification of AGN in Fig. 2 corresponds to a variation in the degree of collimation, going from the winds of Seyferts, to the jets from FR I and then to the powerful jets from FR II type of AGN. Thus we need a criterion for collimation to get a quantitative information on how much the flow will expand. In fact several models have found a 'fastness parameter' α given by

$$\alpha^2 = \frac{L\Omega}{V_a^2}\,, \tag{1}$$

where $L\Omega$ is again the energy of the magnetic rotator and V_a the poloidal Alfvén velocity at the Alfvén transition where $\varpi = \varpi_a$. This parameter was originally introduced by Michel [67] to measure the fastness of the magnetic rotator which accelerates the flow of a cold plasma. In an equatorial wind, when this energy dominates we have a fast magnetic rotator and the wind is magneto-centrifugally driven. Conversely, when thermal acceleration is dominant the magnetic rotator is termed slow. It appeared that α also controls the degree of collimation in several analytical models (e.g. Ferreira [40], Léry et al. [58]). In the numerical approach followed by Bogovalov & Tsinganos [16,96] the degree of collimation is determined by a similar parameter α expressing the ratio of the angular velocity times the Alfvén spherical distance to the initial constant speed of an initially nonrotating split-monopole type of a magnetosphere.

In fact in most of these models boundary conditions were exactly spherically symmetric on the source except for rotation [58,16] or, the magnetocentrifugal forces were dominant in collimating and accelerating the flow [40]. If gas pressure gets important and/or the boundary conditions in density are not spherically symmetric, there seems to be some changes in the degree of collimation, although the role of the fastness parameter remains qualitatively the same [96,24]. In the meridionally self-similar approach followed by Tsinganos et al. [84,94,103,85] the degree of collimation is not only related to the fastness parameter but also to the distribution of the thermal content. In particular, specific criteria for the collimation of winds were derived in the frame of these models [84,85] that we shall summarize here.

Usually in an outflow the thermal input in the form of internal energy and external heating is not all fully converted into other energy forms: unless the terminal temperature is zero, there remains some asymptotic thermal content, $h(\infty, A)$. By subtracting from the total energy $E(A)$ the heat content at infinity, $h(\infty, A)$, we obtain a new streamline constant, $\tilde{E}(A)$, which will be the total convertable specific energy along the given streamline A, i.e., the energy which

can be converted to other forms. Finally, the volumetric total convertable energy is $\rho(r, A)\tilde{E}(A)$.

It turns out that in meridionally self-similar flows the difference of the volumetric convertable energy between a nonpolar streamline and a polar streamline normalized to the volumetric energy of the magnetic rotator $\rho(r, A)L(A)\Omega(A)$ is a constant ϵ' [85],

$$\epsilon' = \frac{\rho(r, A)\tilde{E}(A) - \rho(r, \mathrm{pole})\tilde{E}(\mathrm{pole})}{\rho(r, A)L(A)\Omega(A)} \,. \tag{2}$$

This quantity ϵ' plays a crucial role in the asymptotic shape of the streamlines, in the sense that the necessary condition for cylindrical asymptotics is $\epsilon' > 0$. In other words, cylindrical collimation is controlled by a single parameter, ϵ', representing the variation between a fieldline A and the pole, of the sum of the (volumetric) poloidal and toroidal kinetic energies, the Poynting flux, the gravitational potential and the converted thermal content.

If everything is homogeneous in the medium except rotation and the energy of the magnetic rotator, which always increases away from the axis, then this parameter is very similar to the fastness parameter ($\epsilon' \sim \alpha^2$). Conversely, if the density increases also substantially away from the axis while temperature drops so much that the thermal heating cannot lift the plasma, acceleration will be done at the expense of the Poynting flux. Thus the Poynting flux won't be any longer available to collimate the flow and the degree of collimation will decrease.

In the specific model we are describing, this can be put in a more quantitative form because this parameter splits into two terms [85]:

$$\epsilon' \equiv \mu + \epsilon \,. \tag{3}$$

μ represents the variation across the streamlines of the thermal content that is finally converted into kinetic energy and gives a measure of the thermal pressure efficiency to collimate the outflow. ϵ is the efficiency of the magnetic rotator to collimate.

In fact μ can be written

$$\mu = \frac{P(r, A) - P(r, \mathrm{pole})}{P(r, \mathrm{pole})} \frac{V_\infty^2}{V_*^2} \,, \tag{4}$$

where V_∞ and V_* are the polar asymptotic and Alfvén speeds, and $P(r, A)$ the pressure along the streamline A. For under-pressured flows ($\mu > 0$) the pressure gradient force is outwards helping collimation. Conversely, over-pressured jets ($\mu < 0$) and iso-pressured jets ($\mu = 0$) can collimate only magnetically.

On the other hand, the parameter ϵ is equal to the excess of the magnetorotational energy on a nonpolar streamline which is not used to drive the flow, in units of the energy of the magnetic rotator. It can be evaluated at the base of the flow r_o,

$$\epsilon = \frac{L\Omega - E_{\mathrm{R},o} + \Delta E_{\mathrm{G}}^*}{E_{\mathrm{MR}}} \,. \tag{5}$$

with

$$\Delta E^*_{\rm G} = -\frac{\mathcal{GM}}{r_o}\left[1 - \frac{T_o(\alpha)}{T_o(\rm pole)}\right], \tag{6}$$

where $\mathcal{G}$ is the gravitational constant, $\mathcal{M}$ the central mass and T_o the temperature at the base of the flow r_o. The energy of the magnetic rotator ΩL is mainly stored in the form of Poynting flux, i.e. $E_{{\rm R},o}$ (the rotational energy) is usually a negligible quantity in the above expression. In other words, ϵ measures how much of the energy of the magnetic rotator is not used to escape the gravitational well and is available for magnetic collimation alone. If there is an excess of this energy on non polar streamlines, magnetic forces can collimate the wind into a jet. Thus, when $\epsilon > 0$ we have an *Efficient Magnetic Rotator (EMR)* to magnetically collimate the outflow into a jet, and an *Inefficient Magnetic Rotator (IMR)* if $\epsilon \leq 0$ [85].

Results are summarized in Fig. 8 in the parameter space with typical solutions represented in the poloidal plane. For each solution, the lines are simply a cut in this plane of the magnetic flux tubes, i.e. a projection in this plane of the wounded streamlines as in Fig. 4. First we see that jets from EMR are very well collimated independently of being under- or over- pressured, as illustrated with the solution in Fig. 8a. For IMR the situation is more complex. If the flow is iso- or over-pressured it cannot collimate so it is conical with an asymptotic vanishing pressure as shown in Fig. 8c. Even if it is under-pressured at the base but with vanishing pressure or becoming over-pressured asymptotically as shown in Fig. 8b the cylindrical collimation is very loose because it is due only to the presence of a weak magnetic field. Conversely, if the pressure remains strong all the way it can refocalize strongly the jet and squize it as in Fig. 8d). In fact this last situation looks like a stopped jet somehow.

5.2 Application to the Classification

We found that the asymptotic morphology of the outflow is controlled by the efficiency of the magnetic rotator and the pressure gradient across the streamlines, Eq. (3). The same model shows interesting jet solutions in relation with the various flows seen in AGN (Fig. 8), so we can try to use it in understanding their taxonomy. The efficiency of the magnetic rotator is related to the magnetic properties of the central object in the AGN and/or its disk while the pressure gradient is related to the pressure variation across the streamlines. We may assume that this can be somehow related to the environment through which the jet propagates (this is of course an extra assumption to the model itself). Thus, we may discuss the following possibilities in the framework of the classification scheme of AGN shown in Fig. 2, as they are summarized in Table 1 (this scheme is also shown on the top of Fig 8). In this classification, we move from Type 2 to Type 0 because of orientation effects, as we already discussed. The new interesting element is that a classification from one class to another results now as the efficiency of the magnetic rotator and the environment change.

(I). *Inefficient magnetic rotators, $\epsilon \lesssim 0$, corresponding to radio-quiet AGN (Seyferts, etc...) with uncollimated or, loosely collimated outflows.* One possi-

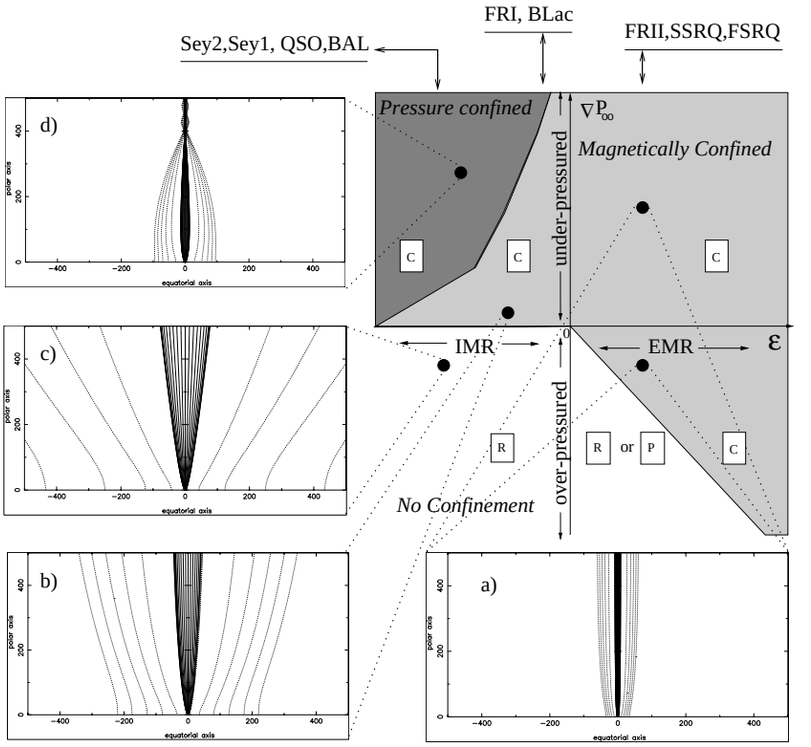

Fig. 8. Degree of collimation obtained as a function of the asymptotic transverse pressure gradient (vertical axis) and the efficiency of the magnetic rotator (horizontal axis) with typical solutions for winds in (b) and (c), jets in (a) and stopped jets in (d). EMR are on the left, IMR on the right, underpressured jets on top and overpressured flows below. "R" corresponds to the domain or radial asymptotics, "P" to paraboloidal ones and "C" to cylindrical. See text for details. After Sauty, Tsinganos and Trussoni [85].

Table 1. AGN classification according to orientation and efficiency of magnetic rotator

Radio-emission (Type: 2, 1, 0)	Magnetic Rotator, ϵ	Collimation, ϵ'
Quiet (Sey.2, Sey.1, BAL & QSO)	Inefficient, $\epsilon \ll 0$	Weak, $\epsilon' \lesssim 0$
Loud (FR I, BL Lac)	Intermediate effic., $\epsilon \sim 0$	Good, $\epsilon' \gtrsim 0$
Loud (FR II, BLRG & SSRQ, FSRQ)	Efficient, $\epsilon > 0$	Tight, $\epsilon' > 0$

bility is that $\epsilon' < 0$ such that the AGN produces a radially expanding outflow (Fig. 8c). This may happen if the source is in a rich environment such that latitudinally we have an over-pressured outflow $\mu > 0$. The other possibility is that $\epsilon' \gtrsim 0$, i.e. ϵ' is marginally positive such that the AGN produces a 'weakly'

collimated jet, i.e., collimation occurs slowly at large distances (Fig. 8b). This may happen, for instance, if we have a latitudinally under-pressured outflow, $\mu > 0$. Hence, if the central source is an IMR ($\epsilon \lesssim 0$) the density drops quite rapidly with the radial distance and this could be related to the weaker outflows in Seyfert 1 and 2 galaxies and radio-quiet QSO's. We cannot exclude that in this case if pressure does not drop rapidly enough we have a stopped jet like in (Fig. 8d) with a shock at the refocalizing point, as it has been suggested for Seyferts (e.g. [82]).

(II). *Efficient magnetic rotators, $\epsilon > 0$, corresponding to radio-loud AGN of high luminosity with well collimated and powerful jets (FRII, etc...).* In this case, since ϵ obtains high positive values, we have a tightly collimated jet (Fig. 8a), regardless of the value of μ, i.e., regardless if the jet propagates in a rich or poor environment.

(III). *Intermediate efficiency magnetic rotators, $\epsilon' > 0$, corresponding to radio-loud AGN of low luminosity (FRI, etc...) with collimated jets.* The two possibilities are either that ϵ is marginally positive and $\mu < 0$, or, ϵ is marginally negative and $\mu > 0$. In this case, we always have $\epsilon' > 0$ and hence the outflows always have asymptotically cylindrical flux tubes. Note also that many extragalactic jets, as deduced from X-ray data on the hot surrounding plasma, seem to be propagating in rich environments [39]. For example, this seems to be the case with FRI type of Radio Galaxies [80].

If the strength of the magnetic rotator reduces, one expects a smooth transition from a jet to a loosely collimated wind and finally to a radial wind. This would correspond to moving from the radio-loud quasars and Blazars of Fig. 2 to the radio-quiet Seyfert galaxies and QSO's. The same transition would be true if the closeby environment, possibly the corona of the central engine, becomes more and more dense. This is also consistent with the different kinds of parent galaxies: early-type (with very low density interstellar gas) for radio-loud AGN, Blazars, QSO, and spiral galaxies for Seyferts.

Despite that the model on which our conclusions are based is clearly nonrelativistic, we conjecture that its basic trends should be preserved in relativistic cases as well if we rely on previous relativistic extensions of non relativistic results [33]. However, collimation of relativistic winds from a spherical source (see [16]) seems rather difficult to achieve because the plasma has a very high effective density in this case (something that somehow fits into the criterion for collimation we previously discussed). This can be solved in two ways. Either, the jet is launched almost along its rotational axis as it is usually done in numerical simulations of relativistic disk winds, or, there is in fact an indication that the relativistic pair plasma beam is confined by an heavier more extended hadronic component which is not relativistic or only mildly so (cf. [46]).

6 Concluding Remarks

In this review we started with a long catalogue of the taxonomy of outflows from AGN and μQuasars but soon we realized that all such outflows share common characteristics that can be understood in physical terms.

First, magnetic collimation of winds into jets appears to be a rather general property of the MHD equations governing plasma outflows (the hoop stress paradigm) and is likely to survive the instabilities of the toroidal magnetic field, although such instabilities indeed must be present and can deeply modify the morphology of the outflow. Pressure gradients also contribute to confine the outflows in addition to toroidal magnetic fields. In extragalactic jets and on scales of several kpc, pressure confinement by the environment seems to be present especially in the less luminous and less collimated ones, while close to the center the jet may be either magnetically or thermally confined.

Second, the transformation of magnetocentrifugal energy into kinetic energy seems to be a natural driving mechanism for outflows from black hole magnetospheres and accretion disks, but the presence of very hot coronae in AGN and the necessity to have a thermal driving along the rotational axis indicates that the contribution from the thermal energy is essential, if not dominant, with appropriate heating processes occurring in the plasma.

And third, the jet composition is likely to be made of both electron/positron pairs and electrons/protons, the first being more likely to be extracted from the central magnetosphere while the second from the more extended corona and surrounding disk.

The MHD acceleration/collimation mechanisms can work in very different astrophysical scenarios, whenever we have a rotating magnetized body such as a supermassive black hole surrounded by an accretion disk. Even though the basic thermal/magnetic driving and confining mechanisms discussed here should be qualitatively valid also for relativistic velocities, more detailed modeling of such relativistic jets from AGN is needed at this point. However a consistent modeling of jets in AGN requires the crossing of all MHD singularities. And we have seen how difficult it is to solve the steady equations and how important it is to put the boundary conditions in numerical simulations carefully. At this point the analogy of horizon/limiting characteristics and ergosphere/elliptic-hyperbolic-transitions may be of some help in the future as there is already a long experience in numerical simulations of black holes and how to tackle with this difficulty.

Finally, we have reviewed the standard unification scheme of AGN. The fact that some classes of objects transform into other classes with the viewing angle seems to be basically secure by now. And, in this article we added that there is a physical criterium separating the various classes among themselves. We see that the degree of collimation does not depend only on the spin of the central black hole or the fueling or the composition of the environment, but in a subtle composition of all these processes which can be expressed in terms of the energetic distribution. In this sense it allows to reconcile the different scenarios proposed to explain the taxonomy of winds and jets around compact objects, in particular the FRII/FRI dichotomy.

References

1. Antonucci, R., *ARA&A*, **31**, 473-521 (1993)
2. Bardeen, J. M., Berger, B.K., *ApJ*, **221**, 105-113 (1978)
3. Baum, S.A., Zirbel, E.I., O'Dea, C.P., *ApJ*, **451**, 88-99 (1995)
4. Begelman, M.C., *ApJ*, **493**, 291-300 (1998)
5. Beskin, V. S., Okamoto, I., *MNRAS*, **313**, 445-453 (2000)
6. Blandford, R.D., in *Active Galactic Nuclei*, eds. T.J.-L. Courvoiser and M. Mayor, Springer, Berlin, pp. 161-275 (1990)
7. Blandford, R. D., Begelman, M.C., *MNRAS*, **303**, L1-L5 (1999)
8. Blandford, R. D., Levinson, A., *ApJ*, **441**, 79-95 (1995)
9. Blandford, R. D., in *Astrophysical Discs* - An EC Summer School, Astronomical Society of the Pacific, Conference series Vol. 160, eds. J. A. Sellwood and J. Goodman, p. 265., AstroPh 9902001 (1999)
10. Blandford, R. D., in *Magnetic Activity in Stars, Discs and Quasars*, eds. D. Lynden-Bell, E. R. Priest and N. O. Weiss., *Phil. Trans. Roy. Soc. A*, Astro-ph 0001499 (2000)
11. Blandford, R.D., Payne, D.G., *MNRAS*, **199**, 883-903 (1982)
12. Blandford, R.D., Rees, M.J., *MNRAS*, **169**, 395-415 (1974)
13. Blandford, R.D., Znajek, R.L., *MNRAS*, **179**, 433-456 (1977)
14. Bodo, G., Rossi, P., Massaglia, S., Ferrari, A., Malagoli, A., Rosner, R., *A&A*, **333**, 1117-1129 (1998)
15. Bogovalov, S.V., *A&A*, **323**, 634-643 (1997)
16. Bogovalov, S.V., Tsinganos, K., *MNRAS*, **305**, 211-224 (1999)
17. Breitmoser, E., Camenzind, M., *A&A*, **363**, 207-225 (2000)
18. Brunthaler, A., et al., *A&A*, **357**, L45-L48 (2000)
19. Camenzind, M., in *Solar and Astrophysical Magnetohydrodynamic Flows*, ed. K.C. Tsinganos, Kluwer (Dordrecht), p. 699-725 (1996)
20. Camenzind, M., in *Astrophysical Jets: Open Problems*, eds. G. Bodo, and S. Massaglia, Gordon and Breach Science Pubs., pp. 3-29 (1998)
21. Capetti, A., Trussoni, E., Celotti, A., Feretti, L., Chiaberge, M., *MNRAS*, **318**, 493-500 (2000)
22. Carter, B., in *Black Holes*, eds. C. DeWitt, and B.S. DeWitt, Gordon and Breach Science Pubs., pp. 57-214 (1972)
23. Casse, F ., Ferreira, J., *A&A*, **353**, 1115-1128 (2000a)
24. Casse, F ., Ferreira, J., *A&A*, **361**, 1178-1190 (2000b)
25. Celotti, A., in *Astrophysical Jets: Open Problems*, eds. G. Bodo, and S. Massaglia, Gordon and Breach Science Pubs., pp. 79-85 (1998)
26. Chan, K.L., Henriksen, R.H., *ApJ*, **241**, 534-551 (1980)
27. Cheng, A.Y.S., O'Dell, S.L., *ApJ*, **251**, L49-L54 (1981)
28. Chiaberge, M., Celotti, A., Capetti, A., *A&A*, **349**, 77-87 (1999)
29. Chiaberge, M., Celotti, A., Capetti, A., Ghisellini, G., *A&A*, **358**, 104-112 (2000)
30. Contopoulos, J., *ApJ*, **450**, 616-627 (1995)
31. Contopoulos, J., Lovelace, R.V.E., *ApJ*, **429**, 139-152 (1994)
32. Corbel, S., Fender, R.P., Tzioumis, A.K., Nowak, M., McIntyre, V., Durouchoux, P., Sood, R., *A&A*, **359**, 251-268 (2000)
33. Chiueh, T., Li, Z-Y., Begelman, M.C., *ApJ*, **377**, 462-466 (1991)
34. Das, T.K., *MNRAS*, **318**, 294-302 (2000)
35. Dhawan, V., Mirabel, I.F., Rodríguez, L.F., *ApJ*, **543**, 373-385 (2000)
36. Fabian, A. C., Rees, M. J., *MNRAS*, **277**, L55-L58 (1995)

37. Fendt, C., Camenzind, M., Appl, S., *A&A*, **300**, 791-807 (1995)
38. Fender, R.P., in "Astrophysics and Cosmology : a collection of critical thoughts", springer Lecture notes in Physics, (2000) astro-ph9907050
39. Feretti, L., Fanti, R., Parma, P., Massaglia, S., Trussoni, E., Brinkmann, W., *A&A*, **298**, 699-710 (1995)
40. Ferreira, J., *A&A*, **319**, 340-359 (1997)
41. Ghisellini, G., Bodo, G., Trussoni, E., Rees, M.J., *ApJ*, **362**, L1-L4 (1990)
42. Ghisellini, G., Bodo, G., Trussoni, E., *ApJ*, **401**, 87-98 (1992)
43. Gopal-Krishna, Wiita, P.J., *A&A*, **363**, 507-516 (2000)
44. Hanasz M., Sol H., Sauty C., *MNRAS*, **316 (3)**, 494-506 (2000)
45. Heinz, J., Begelman M.C., *ApJ*, **535**, 104-111 (2000)
46. Henri, G., Pelletier, G., *ApJ*, **383**, L7-L10 (1991)
47. Heyvaerts, J., Norman, C.A., *ApJ*, **347**, 1055-1081 (1989)
48. Kim, W.-T., Ostriker, E.C., *ApJ*, **540**, 372-403 (2000)
49. Krasnopolsky, R., Li, Z.-Y., Blandford, R., *ApJ*, **526**, 631-642 (1999)
50. Kersalé, E., Longaretti, P.-Y., Pelletier, G., *A&A*, **363**, 1166-1176 (2000)
51. Keppens, R., Goedbloed, J.P., *ApJ*, **530**, 1036-1048 (2000)
52. Koide, S., Shibata, K. Kudoh, T., *ApJ*, **495**, L63-L66 (2000)
53. Königl, A., Kartje, J.F., *ApJ*, **434**, 446-467 (1994)
54. Königl, A., Wardle, M., *MNRAS*, **279**, L61-L64 (1996)
55. Koide, S., Meier, D.L., Shibata, K., *ApJ*, **536**, 668-674 (2000)
56. Kudoh, T., Matsumoto, R., Shibata, K., *ApJ*, **508**, 186-199 (1998)
57. Lery, T., Henriksen R. N., Fiege J., *A&A*, **350**, 254-274 (1999)
58. Lery, T., Heyvaerts, J., Appl, S., Norman, C.A., *A&A*, **347**, 1055-1068 (1998)
59. Lery, T., Frank, A., *ApJ*, **533**, 897-910 (2000)
60. Lery, T., Baty, H., Appl, S., *A&A*, **355**, 1201-1208
61. Li Z-Y., Chiueh, T., Begelman, M.C., *ApJ*, **394**, 459-471 (1992)
62. Li, Z-Y., *ApJ*, **444**, 848-860 (1995)
63. Li, Z-Y., *ApJ*, **465**, 855-868 (1996)
64. Livio, M., *PhR*, **311**, 225-251 (1999)
65. Livio, M., Ogilvie, G.I., Pringle, J.E., *ApJ*, **512**, 100-104 (1999)
66. Lister, M.L., Smith, P.S., *ApJ*, **541**, 66-87 (2000)
67. Michel, F.C., *ApJ*, **158**, 727-738 (1969)
68. Mirabel, I.F., Rodríguez, L.F., *ARA&A*, **37**, 409-443 (1999)
69. Nitta, S-Y., *MNRAS*, **284**, 899-910 (1997)
70. Nobuta, K., Hanawa, T., *ApJ*, **510**, 614-630 (1999)
71. O'Dell, S.L., *ApJ*, **243**, L147-L149 (1981)
72. Okamoto, I., *MNRAS*, **307**, 253-278 (1999)
73. Okamoto, I., *MNRAS*, **318**, 250-262 (2000)
74. Ostriker, E.C., *ApJ*, **486**, 291-306 (1997)
75. Ouyed, R., Pudritz, R.E., *ApJ*, **482**, 712-732 (1997)
76. Parker, E.N., *Interplanetary Dynamical Processes*, Interscience, New York (1963)
77. Pietrini, P., Torricelli-Ciamponi, G., *A&A*, **363**, 455-475 (2000)
78. Pelletier, G., Sol, H., *MNRAS*, **254**, 635-646 (1992)
79. Proga, D., Stone, J.M., Kallman, T.R., *ApJ*, **543**, 686-696 (2000)
80. Prestage, R.M., Peacock, J.A., *MNRAS*, **230**, 131-160 (1998)
81. Rees, M., Begelman, M., Blandford, R., Phinney, E., *Nature* **295**, 17-21 (1982)
82. Rosso, F., PhD thesis, Université de Grenoble (1994)
83. Sakurai, T., *A&A*, **152**, 121-129 (1985)
84. Sauty, C., Tsinganos, K., *A&A*, **287**, 893-926 (1994)

85. Sauty, C., Tsinganos, K., Trussoni, E., *A&A*, **348**, 327-349 (1999)
86. Schatzman, E., *Ann. Astroph.*, **25**, 18-29 (1962)
87. Shibata, K. Uchida, Y., *PASJ*, **42**, 39-67 (1990)
88. Sol, H., Pelletier, G., *MNRAS*, **237**, 411-429 (1989)
89. Sikora, M., Madejski, G., *ApJ*, **534**, 109-113 (2000)
90. Smith, E.P., Heckman, T.M., Bothum, G.D., Romanishin, W., Balick, B., *ApJ*, **306**, 64-89 (1986)
91. Spruit, H.C., Cao, X., *A&A*, **287**, 80-86 (1994)
92. Spruit, H.C., in "Evolutionary Processes in Binary Stars", eds. R.A.M.J. Wijers et al., Kluwer, Netherlands, pp. 249-286 (1996)
93. Tagger, M., Pellat, R., *A&A*, **349**, 1003-1016 (1999)
94. Trussoni, E., Sauty, C., Tsinganos, K., *A&A*, **325**, 1114 (1997)
95. Tsinganos, K., Sauty, C., Surlantzis, G., Trussoni, E., Contopoulos, J., *MNRAS*, **283**, 811-820 (1996)
96. Tsinganos, K., Bogovalov, S.V., *A&A*, **356**, 989-1002 (2000)
97. Uchida, Y., Shibata, K., *PASJ*, **37**, 515-535 (1985)
98. Urry, C.M., Padovani, P., *PASP*, **107**, 803-845 (1999)
99. Usmanov, A.V., Goldstein, M.L., Besser, B.P., Fritzer, J.M., *JGR*, **105**(A6), 12,675-12,695 (2000)
100. Ustyugova, G.V., Koldova, A.V., Romanova, M.M., Chechetkin, V.M., Lovelace, R.V.E., *ApJ*, **516**, 221-235 (1999)
101. Ustyugova, G.V., Lovelace, R.V.E., Romanova, M.M., Li, H., Colgate, S.A., *ApJ*, **541**, L21-L24 (2000)
102. Vlahakis, N., PhD thesis, University of Crete (1998)
103. Vlahakis, N., Tsinganos, K., *MNRAS*, **298**, 777-789 (1998)
104. Vlahakis, N., Tsinganos, K., Sauty, C., Trussoni, E., *MNRAS*, **318**, 417-428 (2000)
105. Wilson, A.S., Colbert, E.J.M., *ApJ*, **438**, 62-71 (1995)

The Evolution of Classical Double Radio Galaxies

Katherine M. Blundell

Oxford University, Astrophysics, Keble Road, Oxford, OX1 3RH, UK

Abstract. Being immensely powerful, and hence detectable out to great distances, classical double radio galaxies have long been recognised as cosmological probes of great potential. Before this potential can be realised, it is necessary to understand the physical mechanisms by which these objects evolve and change with time. This chapter describes how to deduce from classical double radio source *observables* (luminosity, spectral index, redshift and linear size) the essential nature of how these objects evolve and the true relationships between the underlying *physical parameters* (jet-power, age etc). I discuss the key role played by hotspots in governing the energy distribution of the lobes they feed, and subsequent spectral evolution.

1 Unavoidable Problems...

Figure 1 is a cartoon illustration of the nature of the problem to be solved in deducing how radio galaxies change with time. The cone shape represents our light-cone while the short arrows represent the timelines of individual radio galaxies whose lifetimes are short (probably a few 10^8 yrs) compared with the Hubble Time (probably a few 10^{10} yrs) though rather long compared with observational timescales ($\lesssim 10^2$ yrs). One key problem is that we cannot observe any given radio galaxy at well-separated intervals throughout its lifetime and hence we cannot observe how its different physical parameters change with time. We can, however, observe a whole sample of radio galaxies all of which have intercepted our light-cone at different, random points in their lifetimes. After accounting for relevant selection functions, one can then disentangle how the observed distribution of, for example, the luminosities and physical (linear) sizes of the radio galaxies in the sample, having different ages and jet-powers, maps to the underlying relationship between luminosity and linear size in an individual radio galaxy as a function of jet-power and age.

1.1 ...and Selection Functions

One such selection function is determined by the *finding-frequency* of the survey. Different frequencies even within the radio regime, for example 151 MHz and 1.4 GHz, probe different populations as a result of rather different selectivities: added complexities come from the fact that across different radio-frequency regimes different physical processes are responsible for the emission.

At low frequency (e.g. 151 MHz rest-frame) the observed radiation from classical double radio galaxies is dominated by emission from their lobes, and this

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 71–87, 2002.

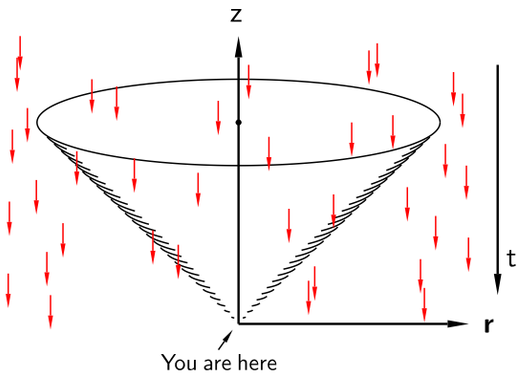

Fig. 1. A schematic illustration of our light-cone. Each arrow represents the timeline of a radio source, whose lifetime is short compared to the Hubble time. Only those arrows which intercept our light-cone are those which we can observe. The point in a radio galaxy's lifetime when it is intercepted by our light-cone is of course random.

emission is largely isotropic. In GHz regimes a rather higher fraction of the luminosity comes from the cores and the hotspots. When fast-moving jet-material is moving along a direction close to our line-of-sight the phenomenon of Doppler boosting enhances this emission very significantly and in well-understood ways (Begelman, Blandford & Rees 1984). In this way a GHz finding-frequency can preferentially select objects with a small angle to the light-of-sight which may intrinsically have low jet-powers. In addition with GHz surveys very young sources such as GPSs/CSOs may be found which are often synchrotron self-absorbed or free-free absorbed at lower frequencies.

Unfortunately, the effect of a single finding-frequency varies as a function of the redshift of the sources being surveyed. For a given survey finding-frequency $\nu_{\rm obs}$, the emission of a particular radio galaxy at redshift z is sampled at rest-frame $(1+z)\times\nu_{\rm obs}$. So for example a finding-frequency of 330 MHz samples an object at redshift $z=5$ at a frequency of over 1 GHz in its rest-frame.

These are the first hints that a radio survey, even one with a low finding-frequency, even when completely identified with spectroscopically determined redshifts, does not constitute a 'randomly-drawn, uniform sample' of classical double radio sources over all redshifts.

2 What Are Radio Galaxy Observables?

The primary observables for a given radio source are, for the purposes of this chapter, to be regarded as (1) the **redshift** (z) of the active nucleus associated with the radio structure, (2) the physical or **linear size** (D) of the radio structure [derivable from the redshift and the direct measurement of the angular separation of the hotspots at the outermost edges of the radio lobes, in some chosen cosmological model], (3) the **spectral shape** (α) of the radio emission

(derivable from measurements of the flux density of the source over a range of radio frequencies, see §3.1) and (4) the **luminosity** (P) at some common rest-frame frequency derived from the spectral shape and the redshift in the same chosen cosmological model. This gives a 4-dimensional parameter space of observables (see Fig. 2) from which we have to determine the relationships between the underlying physical parameters of radio sources. These physical parameters include the bulk kinetic power transported by the jets, and the age of the radio source as we observe it. In the 4-D space, we have 6 pairs of observables whose interdependences (or lack, thereof) have to be considered. After some considerations about the correct way to measure each of these observables, we look at each of these six dependences in turn.

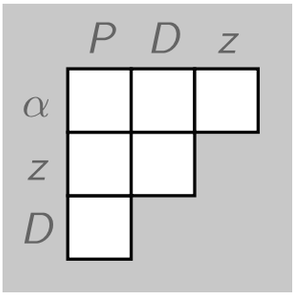

Fig. 2. The 4-dimensional P, D, z, α space of the observables. In this space, we have 6 pairs of observables whose interdependences (or lack, thereof) have to be considered.

3 Measurement of Source Properties

3.1 Spectral Index

The integrated spectra of classical double radio sources are frequently observed to be curved, in the concave sense (see Fig. 3). Rather more rarely they are observed to be power-laws, though surprisingly frequently this is the assumed form of the spectrum used in radio source modelling in the past.

In order to properly measure the spectral index at a common rest-frame frequency, it is necessary to fit the curvature of the spectra and perform a k-correction to take account of the redshifted sampling of the spectra. All members of the 3C, 6C and 7C complete samples used in the study of Blundell, Rawlings & Willott (1999) had spectral fitting and k-corrections performed using flux-density measurements from the literature for each object at 365 MHz, 408 MHz, 1.4 GHz and 4.86 GHz in addition to measurements from pointed VLA observations as well as from the survey measurements at 151 MHz (and in the case of the 3C sample, lower frequencies still from Laing & Peacock 1980). Hereafter in this chapter I refer to the instantaneous slope of the log flux-density – log frequency

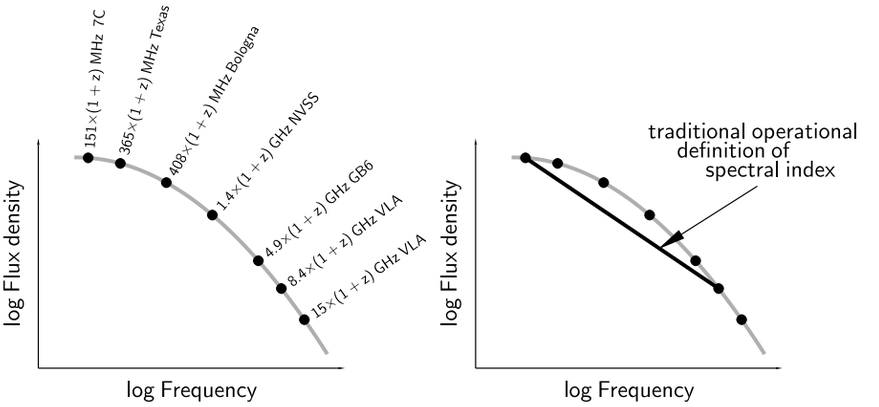

Fig. 3. Figure to illustrate the integrated spectral shapes of radio sources, deduced from the many surveys which are available to us these days. The traditional measurement of spectral index (the gradient of the curve on a log flux-density *v.* log frequency plot) assumes that the spectrum follows a simple power-law. A correct measurement of spectral index involves fitting the curvature of the spectrum and correcting for the redshifting of the emission.

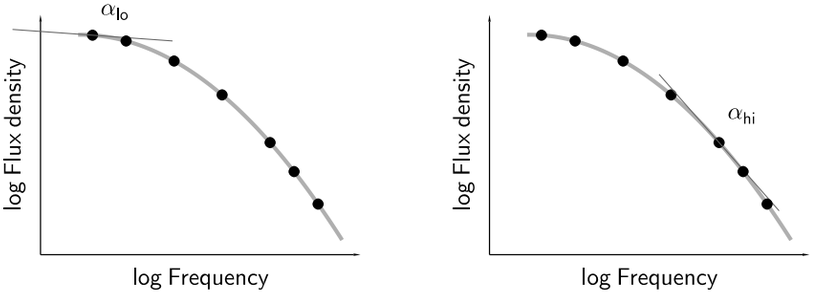

Fig. 4. *Left:* The integrated spectra of radio sources at low rest-frame frequency is dominated by emission from the lobes. *Right:* At high (GHz) rest-frame frequency this emission is contaminated by emission from the hotspots and core. The mapping between the jet-power and the luminosity of the core and the hotspots is less direct than the mapping between the jet-power and the luminosity of the lobes.

plot evaluated at rest-frame 151 MHz as the spectral index at 151 MHz or $\alpha_{\rm lo}$, and evaluated at rest-frame 5 GHz as the spectral index at 5 GHz or $\alpha_{\rm hi}$.

3.2 Luminosity

While it is a good observable, luminosity is not a fundamental property of a radio source in the sense that it is not a direct measure of the rate at which energy is transported by the jets i.e. it is not the underlying jet-power of a radio source. This is because in addition to depending on the details of the environment into which the source is expanding, the luminosity also depends

on the magnetic field distribution and the distribution in Lorentz factor of the synchrotron particles and on the age of the source. This is true even if the bulk kinetic jet-power remains constant. The luminosity also depends on the frequency regime in which it is being evaluated, for example, it is especially but not exclusively true for those objects having significantly Doppler boosted core emission. For the source described here, all luminosities are evaluated at rest-frame 151 MHz, following the spectral fitting procedure mentioned in §3.1.

3.3 Redshift

The ideal way to measure a redshift is by comparing the observed wavelengths of identified emission lines with their rest-frame wavelengths. Many studies of samples of radio sources in the past have required the use of redshift estimators for a significant fraction of the sample (e.g. Dunlop & Peacock 1993). The most common estimator employs the K–band (2.2 μm) magnitude. Use of this estimator is motivated by the small spread in the near-infrared Hubble Diagram (K–z relation) for radio galaxies in the 3C sample (Lilly & Longair 1984; Best, Longair and Röttgering 1998). It is now known that simple extrapolation of the 3C relation to higher redshifts and/or to fainter radio samples (which contain a smaller fraction of radio luminous sources at any given redshift) leads to both large uncertainties in the estimated redshifts, and to systematic biases (Rawlings et al. 1998). For the sources in the 3C, 6C and 7C samples used in the study described here, the redshifts are almost exclusively spectroscopic. In the very few cases where spectroscopic redshifts had not been secured they were estimated by obtaining multi-colour (optical and near-infrared) information and fitting spectral energy distributions (SEDs) with template galaxy spectra (Willott, Rawlings & Blundell 2001).

3.4 Linear Size

Angular sizes of radio sources are measured by taking the sum of the separation from the core to each hotspot or, in the absence of a core, the hotspot-to-hotspot separation. Angular sizes are converted into linear sizes using the redshift of a source and assuming a particular cosmological model (e.g. Carroll, Press & Turner 1992).

4 Minimizing the Effects of Spurious Correlations

4.1 Luminosity and Redshift Dependences

The consequences of a flux-limit, together with the rarity of classical doubles in the local Universe mean that for a single sample, such as 3C, there will be a tight correlation between the luminosities (P) and redshifts (z) of the members of that sample. This is known as the P–z degeneracy. This is illustrated in the left panel of Fig. 5. Developing models of radio sources based on just a single flux-limited sample runs the risk of modelling and misconstruing the P–z degeneracy.

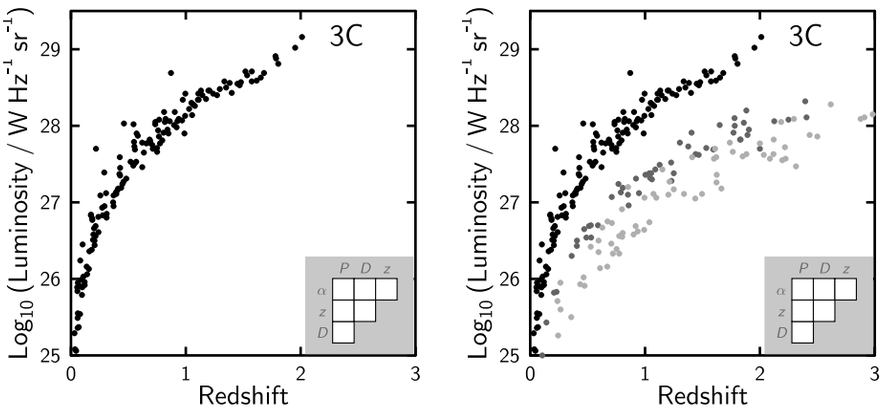

Fig. 5. *Left:* The coverage of the P–z plane by the 3C sample gives a very tight correlation of luminosity and redshift. The lower boundary of the distribution of points is caused by the flux-limit of this survey (originally 10.9 Jy at 178 MHz). *Right:* The study of Blundell, Rawlings & Willott (1999) is based on three complete samples selected at different flux-limits at 151 MHz, a frequency sufficiently low that Doppler biases are unimportant. See discussion on the finite widths of these regions in Blundell, Rawlings & Willott (1999).

To circumvent such a problem, coverage of the P–z plane was increased by using complete samples of radio sources which are selected at lower flux-limits (Rawlings et al. 1998). While the flux-limit of the 3C sample is 12 Jy at 151 MHz, the flux-limit of the faintest sample (7C) is 0.5 Jy. This was a huge and painstaking observational effort but means that for the first time at low frequency coverage of the P–z plane is substantially improved over that obtained from 3C alone as illustrated by the right panel in Fig. 5. This is essential in the quest to deduce the true dependences between the four observables, P, D, z and α.

5 Salient Correlations and Non-correlations

This section describes in general terms the main dependences and their interpretations; no statistical justification is here presented (statistics are not, after all, as interesting as they are important) but a thorough statistical analysis of these and other dependences may be found in Blundell, Rawlings & Willott (1999).

5.1 The D–$\alpha_{\rm lo}$ Dependence

In measuring spectral index in the accurate way described in §3.1, a new dependence between one pair of the $[PDz\alpha]$ observables was discovered. Fig. 6 shows that the greater the linear size of a radio source the steeper is its low frequency spectral index. This was the strongest correlation found out of all of those for the

complete samples studied by Blundell, Rawlings & Willott (1999). This dependence is likely to be a consequence of adiabatic expansion losses and is telling us that the magnetic fields decrease as the lobes expand, for the following reason. (Other possibilities are considered, and rejected, in Blundell, Rawlings & Willott 1999.) For a fixed observing frequency ν, the relationship between the Lorentz factor (γ) of the particles giving rise to that emission and the ambient magnetic field B is given by:

$$\gamma = \left(\frac{m_{\rm e}}{eB} 2\pi\nu\right)^{\frac{1}{2}}, \tag{1}$$

where e is the charge on an electron and $m_{\rm e}$ is its rest-mass. Thus as the magnetic field decreases, higher-γ particles will be responsible for radiation at that fixed observing frequency. Since for lobe emission the spectrum of the energy distribution of particles is steep and frequently curved these higher-γ particles will be from a *steeper* part of the energy distribution, and hence a *higher* rest-frame 151 MHz spectral index will be observed. Thus emission at this particular frequency from the larger (probably older) linear size objects is likely to arise from higher Lorentz factor particles than those in smaller (probably younger) linear size sources. Another, important, way of looking at this is to consider that if as a source increases in age its magnetic field decreases as it expands the Lorentz factors of particles responsible for the emission at some fixed frequency gradually increase. The canonical energy distribution of synchrotron particles (that there are *fewer* as one considers *higher* Lorentz factors) means that as higher Lorentz factor particles provide the emission at the fiducial frequency so the luminosity of the source must decrease since there are *fewer* particles.

5.2 The P–$\alpha_{\rm lo}$ Correlation

A correlation between the luminosities of classical doubles and their spectral indices was first suggested in 1960 by Heeschen, and more recently by Laing & Peacock (1980). We find this correlation to be present for the combination of the 3C, 6C and 7C complete samples in the sense that the more powerful the radio source, the steeper is its low-frequency spectral index. Note however that it is only with the *combination* of complete samples used in the study of Blundell, Rawlings & Willott (1999) which has broken the P–z degeneracy that one can be sure that this truly is a dependence of P and α rather than z and α (the relationship, or lack thereof, between this pair of observables is discussed in §5.3 and 5.4).

One way of depicting the P–α correlation is to show for the P–z plane the effects of low-frequency spectral index on the flux-limits and how sources from the different complete samples cluster around these flux-limits (see Fig. 7).

We have modelled this P–α correlation as coming from a dependence of the steepness of the energy distribution of particles injected into the lobe on jet-power. We do not presume that the injection spectrum is universally given by a power-law with the canonical value of $\alpha = 0.5$. We model the steepening of the energy distribution of particles as being controlled largely by classical

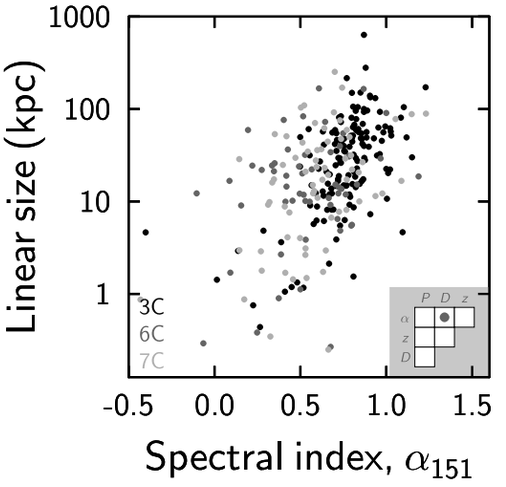

Fig. 6. The linear sizes of the radio sources plotted against their spectral indices evaluated at 151 MHz in the rest-frame. If one focusses just on the dark (3C; Laing, Riley & Longair 1983) points a very tight correlation is seen between this pair of observables. Such a correlation only becomes apparent if the spectral shape is correctly fitted and de-redshifted as described in §3.1.

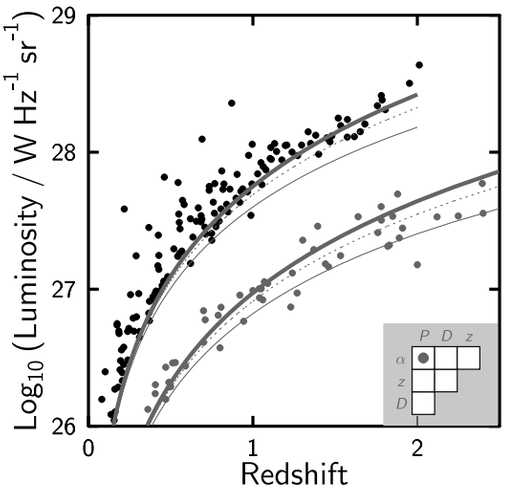

Fig. 7. The luminosities of the radio sources plotted against their redshifts, with the flux-limits for the samples overlaid for different assumed values of the low-frequency spectral index. Only the 3C and 6C samples are plotted; the 7C data are omitted for clarity but may be seen in figure 7 of Blundell, Rawlings & Willott (1999). It can be seen that those 3C sources with $P > 27.5\,\mathrm{W\,Hz^{-1}\,sr^{-1}}$ all avoid the region between the $\alpha = 0.5$ flux-limit and the $\alpha = 0.8$ flux-limit. There is no such avoidance seen for the lower-power 6C sources over the same redshift range.

synchrotron loss suffered in the high magnetic field of the hotspot. We specifically invoke that the hotspot magnetic fields are higher in those sources with higher jet-powers (as detailed in Blundell, Rawlings & Willott 1999). We choose to normalise this dependence of hotspot magnetic field on jet-power by assuming equipartition (of the energy in the radiating particles and the energy in the magnetic field) in the hotspot (for which there is good evidence in the case of Cygnus A, Harris, Carilli & Perley 1994). The significance of a higher magnetic field in the hotspot is that it changes the energy distribution of the plasma which is in the hotspot. If this plasma has its energy distribution steepened in the hotspot it will remain steepened when it is in the lobe. It is interesting in this regard that Dennett-Thorpe et al. (1999) find for a sample of 3C sources that the spectral indices of the hotspots themselves are steeper with increasing redshift (or luminosity equivalently, since this is for a single flux-limited sample) — see their figure 4.

The role of the hotspot is further discussed in §6.

5.3 Redshift and Low-Frequency Spectral Index

The low-frequency spectral index $\alpha_{\rm lo}$ (evaluated at 151 MHz in the rest-frame) is less subject to the effects of synchrotron and inverse Compton losses off the CMB than higher-frequency spectral indices ($\alpha_{\rm hi}$) evaluated in the GHz regime. As such, $\alpha_{\rm lo}$ is potentially a closer indicator of the energy distribution as initially injected than is $\alpha_{\rm hi}$. We find that there is no dependence on redshift of $\alpha_{\rm lo}$, as shown in Fig. 8 (upper panel).

5.4 Redshift and High-Frequency Spectral Index

In contrast with the picture in §5.3 there is a tendency for the GHz spectral index to increase with increasing redshift in Fig. 8 (lower panel). This points to the increasing importance of inverse Compton losses off the CMB with increasing redshift for the lobes, which was first pointed out by Krolik & Chen (1991). Note that this tendency is to some extent offset by the increasing fractional contribution of the core and the hotspots (with their rather different spectra and dominant loss mechanisms) to the luminosity at GHz frequencies.

5.5 The D–z Correlation

The 'linear-size evolution' of radio sources has been known for many years (Kapahi et al. 1987, Barthel & Miley 1988). The combination of complete samples selected at low frequency by Blundell, Rawlings & Willott (1999) shows a decrease of linear size with increasing redshift which is quite mild (see Fig. 9), in the sense that the parameterisation of linear size at constant luminosity and spectral index is given by $D \propto (1+z)^{-0.96}$. We return to the physical mechanism responsible for linear-size evolution of classical double radio sources in §6.

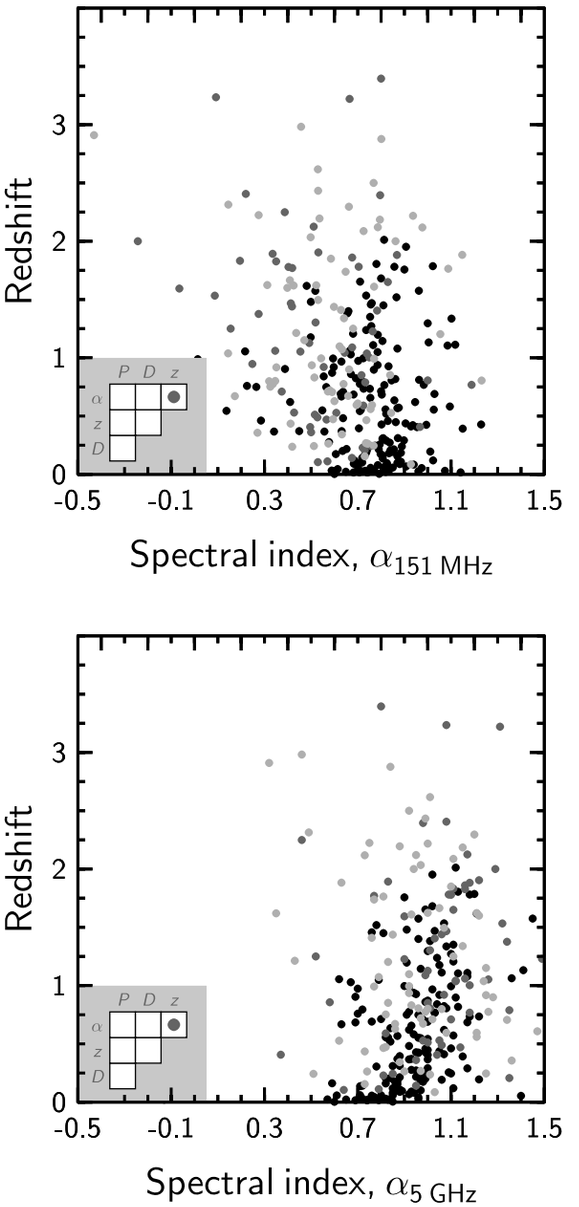

Fig. 8. Upper panel: redshift v spectral index evaluated at 151 MHz in the rest-frame. Lower panel: redshift v spectral index evaluated at 5 GHz in the rest-frame.

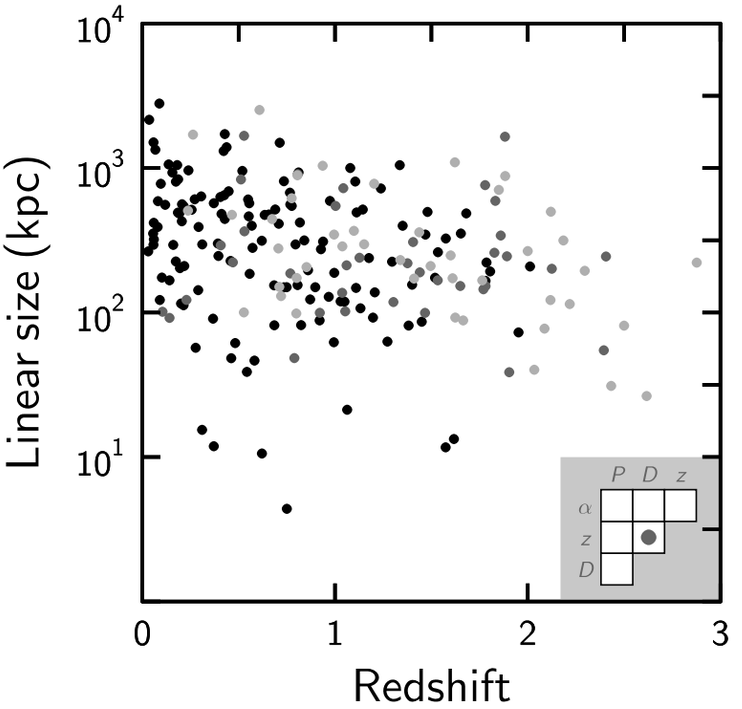

Fig. 9. The linear sizes of the radio sources from the complete samples studied by Blundell, Rawlings & Willott (1999) versus their redshifts.

5.6 The *P*–*D* Plane

The *P*-*D* plane depicted in Fig. 10 shows that there is no strong correlation between the luminosities and linear sizes of radio sources in the combination of complete samples. Given the sampling functions alluded to at the start of this chapter which influence the nature of the objects are selected by our surveys, it should *not* be claimed that the absence of a correlation between P and D in this plot in any way might imply that for any given source as D grows, P will not evolve.

6 Refining Models for Luminosity Evolution

We have formulated a model for evolving radio galaxies which reproduces the above dependencies, and others, and this is detailed in Blundell, Rawlings & Willott (1999). This model is the first model for radio galaxy evolution which incorporates the role played by the hotspot.

A key feature of the model is that we decouple the role played by the 'head' of the radio source (the outermost region of the lobe, see Fig. 11), which drives

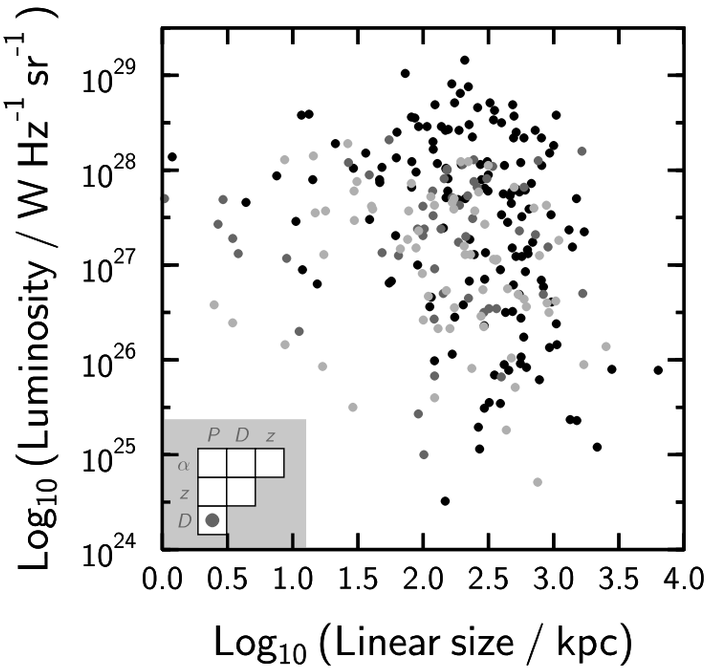

Fig. 10. The luminosities of the radio sources from the three complete samples versus their linear sizes.

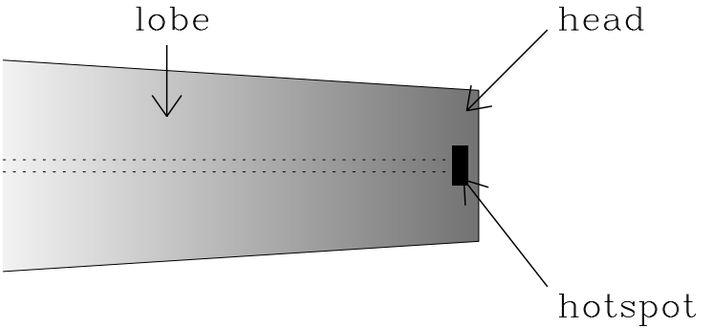

Fig. 11. Illustration of the nomenclature of a radio galaxy as used throughout this chapter.

the expansion of the radio source, from the distinct role played by the 'hotspot' (the compact region of enhanced surface-brightness towards the outer edge of the radio lobe) which is likely to cause significant radiative losses because of its high magnetic field. The role played by the hotspot has two aspects.

First, the magnetic field of the hotspot causes significant radiative losses on synchrotron particles during their dwell-time in the hotspot, in the period of time before they are injected into the lobe. We were led to this as discussed in §5.2 by our observation that the radio sources in our complete samples seemed to strongly indicate that the more luminous a radio-source was, the steeper its *low-frequency spectrum* was (for example, when evaluated at rest-frame 151 MHz). This rest-frame frequency regime is dominated by emission from the lobes, rather than from hotspots and cores, as is the case at GHz frequencies. Modelling a steeper spectrum in a more powerful source came naturally out of invoking stronger magnetic fields (hence stronger radiative losses) in sources with higher jet-powers. By equating the jet-thrust and the pressure in the compact hotspot, and invoking equipartition in the hotspot, we find that the magnetic energy density in the hotspot is proportional to the bulk kinetic power transported in the jet.

Second, as a plasma element expands out of a compact hotspot into a lobe whose pressure is considerably lower (and continues to become lower as the radio lobe expands and gets older) it suffers enhanced adiabatic expansion losses compared to the expansion losses suffered by that element of plasma once it continues to dwell in the lobe. This is consistent with observations of classical doubles which invariably show highly compact hotspots — embedded towards the outermost edges of the smooth, low surface-brightness, extended emission which comprises the lobe — albeit with a bewildering menagerie of shapes and structures (Leahy et al. 1997, Hardcastle et al. 1997 & Black et al. 1992).

Early models (Scheuer 1974; Baldwin 1982) as well as more recent models (Kaiser, Dennett-Thorpe & Alexander 1997; Blundell, Rawlings & Willott 1999) of radio source evolution, which make reasonable assumptions about the environments into which radio sources expand, predict that the luminosity of any individual radio source decreases as time — or age — increases. This luminosity decline is enhanced when the role of the hotspot is included and illustrative 'tracks' across the P–D plane of radio sources with given jet-powers and redshifts are shown in Fig. 12.

A combination of the dramatically declining luminosity-with-age of the high-jet-power sources, their scarcity in the local Universe, together with the harsh reality of survey flux-limits means that very powerful sources with large linear sizes are rarely seen. This resolves the old problem first posed by Baldwin in 1982 that there is a dearth of large *and* powerful classical doubles. This problem had been a puzzle: intuitively [and as later modelled by Falle (1991)] the more powerful radio-sources were expected to expand faster than those with lower jet-powers, so the laws of probability alone suggest it should be easy to find large, powerful sources. But they are not: sources with $P > 10^{27}\,\mathrm{W\,Hz^{-1}\,sr^{-1}}$ and with $D > 1\,\mathrm{Mpc}$ are extremely rare in existing surveys.

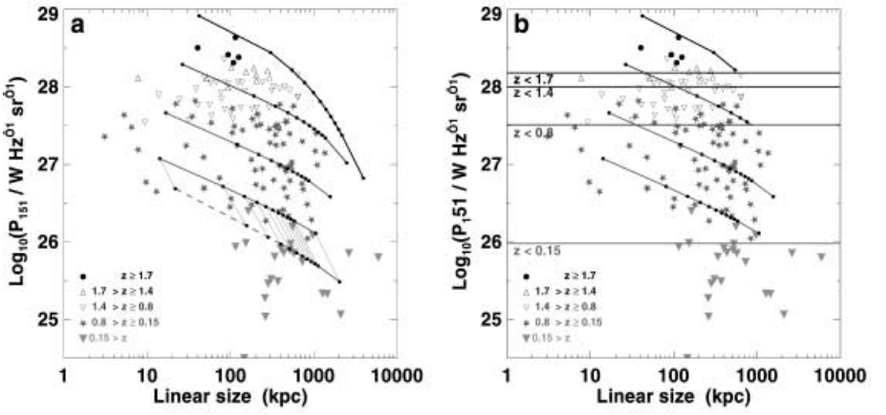

Fig. 12. Overlaid on the 'P–D' plane for the 3C sample in **a** are model tracks tracing out the evolution in luminosity and linear size of four example radio sources, with from top to bottom jet-power, $Q = 5 \times 10^{39}$ W at $z = 2$, $Q = 1 \times 10^{39}$ W at $z = 0.8$, $Q = 2 \times 10^{38}$ W at $z = 0.5$ and $Q = 5 \times 10^{37}$ W at $z = 0.15$. Small dots lie on all tracks to indicate the (rest-frame) ages 1, 10, 20...100, 200 Myr. The dashed line indicates how the lower track luminosity reduces by < half an order of magnitude if the ambient density becomes an order of magnitude lower. In **b** the horizontal lines represent the luminosities at which the 3C flux-limit of 12 Jy takes its effect at the different redshifts indicated. The same tracks as in **a** are plotted but here in **b** they are truncated when they reach the flux-limit. Source expansion in these plots is described by the model of Falle (1991). In his model higher jet-power sources expand more quickly; thus these objects fall through the flux-limit at a younger age. The luminosity evolution is as described by the model of Blundell, Rawlings & Willott 1999, which incorporates the role of the hotspots.

7 The Youth-Redshift Degeneracy

Declining luminosity tracks in combination with the consequences of applying a flux-limit at different redshifts inevitably give rise to an effect termed the 'youth-redshift degeneracy' by Blundell & Rawlings (1999). This is the consequence that in a given flux-limited sample, those sources at higher redshift must be younger than their low redshift counterparts. This effect applies very strongly to powerful sources when the role of the hotspot is incorporated into the model which generates P–D tracks, as described in §6.

The youth-redshift (YZ) degeneracy means it is necessary to be careful about the correct interpretation of 'trends with redshift'. In many cases it is wrong to attribute such trends to changes in *cosmic epoch* when the effects arise because of *source age*. We consider briefly five such examples of effects which occur because the high redshift sources are young sources, and hence are being observed not long after the jet-triggering event (which is possibly a galaxy-galaxy merger, Sanders et al 1988).

Linear-size evolution and the YZ-degeneracy. The linear size evolution observed in low-frequency flux-limited samples of classical double radio-sources (Kapahi et al. 1987, Barthel & Miley 1988, Blundell, Rawlings & Willott 1999) arises because the high–z sources are younger, hence tend to be shorter. It is the positive dependence on jet-power of the rate at which the lobe-lengths grow (Falle 1991) which contributes to the linear size evolution being as mild as it is (see §5.5 and Blundell, Rawlings & Willott 1999).

Structural distortion and the YZ-degeneracy. Barthel and Miley (1988) had suggested that higher redshift environments are denser and more inhomogeneous than at low redshift since they found increased distortion in the structures of their high-z sample of steep-spectrum quasars compared with their low-z sample. Sources which are younger may have the passage of their jets considerably more disrupted where there is a higher density and greater inhomogeneity in an environment which has recently ($< 10^7$ years) witnessed a jet-triggering event [perhaps a galaxy-galaxy merger (Sanders et al 1988)]. A general trend of denser inter-galactic environments at high-z cannot be inferred from their result.

The alignment effect and the YZ-degeneracy. Where the alignment effect of optical light along radio-jet axes is caused by star-formation, it will be more easily triggered close in to the host galaxy or within the product of a recent merger (assuming this is the jet-triggering event) than at distances further out sampled by the head of an expanding radio-source later in its lifetime. Where the alignment effect is caused by dust-scattered quasar light, the certain youthfulness of distant radio-galaxies alleviates the near discrepancy (De Young 1998) between radio-source ages and the time-scale for which dust grains can survive in the presence of shocks caused by the advancing radio-jets. The youth–redshift degeneracy is consistent with the finding that the smallest sources in a sample of $z \sim 1$ radio galaxies (all with very similar luminosities) are those which are most aligned with optical emission (Best, Longair and Röttgering 1996). Best et al. had remarked that the sequence of decreasing optical aligned structure with increasing radio size could be naturally interpreted by comparing it with different phases of the interaction of the radio jets with the interstellar and intergalactic media as the radio-sources get older.

Faraday depolarisation and the YZ degeneracy. Garrington & Conway (1991) found a tendency for depolarisation to be higher in sources at higher z. Objects with higher z which are younger may be in much more recently merged environments with the consequence that inhomogeneities in density or magnetic field will more readily depolarise the synchrotron radiation from the lobes. Moreover, higher-z sources being younger and tending to be somewhat shorter will be closer in to the centre of the potential well. The higher density in this region will enhance the observed depolarisation.

The YZ degeneracy and dust emission. Many of the highest-z radio-galaxies have gas masses comparable to gas-rich spiral galaxies (Dunlop et al. 1994; Hughes et al. 1998; Archibald et al. 2001) and inferred star-formation rates which, in the local Universe, are rivalled only by galaxy-galaxy mergers like Arp 220 (Genzel et al 1998). If high-z objects are being viewed during a similar

merging of sub-components the associated star formation could be responsible for a significant fraction of the stellar mass in the remnant galaxy. Since the high-z radio-galaxies — those which Hughes et al. and Archibald et al. detected with SCUBA — are necessarily young ($< 10^7$ years, see Fig. 12), and since the whole merger must take a few dynamical crossing times, or 10^{8-9} years, the implication is that the event which triggered the jet-producing central engine is synchronised with massive star formation in a gas-rich system, perhaps as material streams towards the minimum of the gravitational potential well of the merging system. The YZ degeneracy may help explain why few lower-z radio-galaxies show similarly large (rest-frame) far-infrared luminosities compared to the high-z population: they are being observed significantly longer after the jet-triggering event.

Acknowledgements

I warmly thank my collaborators Steve Rawlings and Chris Willott. I am very grateful to the Royal Society for a University Research Fellowship.

References

1. E.N. Archibald, J.S. Dunlop, D.H. Hughes, S. Rawlings, S.A. Eales & R.J. Ivison: MNRAS, **323**, 417 (2001)
2. J.E. Baldwin: 'Evolutionary tracks of extended radio sources'. In: *Extragalactic radio sources, 97th IAU Symposium at Albuquerque, New Mexico, USA, August 3–7, 1981*, ed. by D.S. Heeschen & C.M. Wade (Reidel, Dordrecht 1982) pp. 21–24
3. P.D. Barthel & G.K. Miley: Nature, **333**, 319 (1988)
4. M.C. Begelman, R.D. Blandford & M.J. Rees: Reviews of Modern Physics, **56**, 255 (1984)
5. P.N. Best, M.S. Longair, & H.J.A. Röttgering: MNRAS, **280**, 9L (1996)
6. P.N. Best, M.S. Longair, & H.J.A. Röttgering: MNRAS, **295**, 549 (1998)
7. A.R.S. Black, S.A. Baum, J.P. Leahy, R.A. Perley, J.M. Riley, P.A.G. Scheuer: MNRAS, **256**, 186 (1992)
8. K.M. Blundell, S. Rawlings: Nature, **399**, 330 (1999a)
9. K.M. Blundell, S. Rawlings, C.J. Willott: AJ, **117**, 677 (1999b)
10. S.M. Carroll, W. Press, E.L. Turner: ARAA, **30**, 499 (1992)
11. G. Cotter, S. Rawlings, R. Saunders: MNRAS, **281**, 1081 (1996)
12. D.S. De Young: ApJ, **507**, 161 (1998)
13. J.S. Dunlop, J.A. Peacock: MNRAS, **263**, 936 (1993)
14. J.S. Dunlop, D.H. Hughes, S. Rawlings, S.A. Eales, M.J. Ward: Nature, **370**, 347 (1994)
15. S.A.E.G. Falle: MNRAS, **250**, 581 (1991)
16. S.T. Garrington, R.G. Conway: MNRAS, **250**, 198 (1991)
17. R. Genzel, D. Lutz, L. Tacconi: Nature, **395**, 859 (1998)
18. M.J. Hardcastle, P. Alexander, G.G. Pooley, J.M. Riley, MNRAS, **288**, 859 (1997)
19. D.E. Harris, C.L. Carilli, R.A. Perley: Nature, **367**, 713 (1994)
20. D.S. Heeschen: PASP, **72**, 368 (1960)

21. D.H. Hughes, J. S. Dunlop, E. N. Archibald, S. Rawlings, S. A. Eales: in *Cosmological evolution of the submmillimetre luminosity of high-redshift radio galaxies in* The birth of galaxies (ed Guiderdoni, B.) (Frontieres, Paris, 1998).
22. C.R. Kaiser, J. Dennett-Thorpe, P. Alexander, MNRAS, **292**, 723 (1997)
23. V.K. Kapahi, C.R. Subrahmanya, V.K. Kulkarni, J. Astrophys. Astron., **8**, 33 (1987)
24. J.H. Krolik & W. Chen: AJ, **102**, 1659 (1991)
25. R.A. Laing, J.M. Riley & M.S. Longair: MNRAS, **204**, 151 (1983)
26. R.A. Laing & J.A. Peacock, MNRAS, **190**, 903 (1980)
27. J.P. Leahy, A.R.S. Black, J. Dennett-Thorpe, M.J. Hardcastle, S. Komissarov, R.A. Perley, J.M. Riley, P.A.G. Scheuer, MNRAS, **291**, 20 (1997)
28. S.J. Lilly, M.S. Longair: MNRAS, **211**, 833 (1984)
29. D.B. Sanders, B.T. Soifer, J.H. Elias, B.F. Madore, K. Matthews, G. Neugebauer, N.Z. Scoville, ApJ, **325**, 74 (1988)
30. P.A.G. Scheuer: MNRAS, **166**, 513 (1974)
31. C.J. Willott, S. Rawlings & K.M. Blundell: MNRAS, **324**, 1 (2001)

Blazars

Annalisa Celotti

S.I.S.S.A., via Beirut 2-4, I-34014 Trieste, Italy

1 What Are They?

Though a blazar can be characterized by different observational properties (typically flat radio spectrum, variability, linear polarization in the radio and/or optical band...), a less operational but plausibly more physical definition is that blazars are Active Nuclei of Galaxies whose emission is dominated by radiation from non–thermal plasma moving at relativistic speed in a direction close to the line of sight, as proposed over 20 years ago by Blandford & Rees (1978). The relativistic speed implies that the radiation observed is amplified (beamed) in the direction of motion because of aberration, time and energy transformations. This definition of blazars broadly includes BL Lac objects (BL Lac) and Flat Spectrum Radio Quasars (FSRQ), where the former ones are typically featureless while large and intense line emission is detectable in the optical–UV spectrum of quasars.

The relativistic plasma moves along/forms a jet, which represents the channel through which the central engine supplies energy to the extended radio structures, reaching 100 kpc - Mpc distances from the nucleus. While much has been learned – observationally and theoretically – about blazars, key issues are still matter of debate: how is (in which form) energy transported ? how is part of it radiatively dissipated ? how/why are jets formed, collimated and accelerated close to the central engine ? what is their connection with the phenomenon of material accreting onto a black hole ?

In the following I will selectively discuss some of these issues and stress how processes associated to the relativistic flows, which constitute the common denominator of this conference, are at the core of the blazar phenomenon. Apologies for my biases and the incompleteness of such account are certainly owed to many colleagues.

1.1 Emission

The broad band spectral energy distribution (SED) of blazars extends from radio to γ–ray frequencies, and is currently described, in a νL_ν representation, as constituted by two broad bumps, peaking in the mm to soft (and even hard) X–ray and in the gamma-ray (MeV to TeV) bands, respectively.

The detection by the *Compton Gamma Ray Observatory* (e.g. Hartman et al. 1999) and Cherenkov telescopes (e.g. Weekes et al. 1996; Petry et al.

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 88–100, 2002.

1996) of γ–ray emission from blazars is indeed of primary importance for the understanding of these sources. Such intense and highly variable emission in fact represents in most sources the bulk of the radiative output. Furthermore it provides independent evidence for relativistic beaming and sets important constraints on the location of the radiative dissipation along the jet: if it were too compact or close to the accretion disk and its corona, γ–rays would not survive absorption via γ–$\gamma \rightarrow e^{\pm}$ process, and if it were too large the emission could not vary too rapidly (Ghisellini & Madau 1996).

However it has not been possible so far to uniquely pin down the dominant emission mechanism(s). Models can be schematically (and roughly) divided into two broad classes. Within one class it is assumed that bulk motion and/or electromagnetic energy is directly converted into internal energy of relativistic electrons with a non–thermal distribution. These emit radiation via the synchrotron and inverse Compton process, giving origin to the low and high energy spectral components, respectively. The seed photons for the inverse Compton scattering could be both the synchrotron photons themselves (Synchrotron Self Compton, SSC) and photons produced externally to the emitting jet (External Compton, EC), e.g. isotropized nuclear radiation, broad line photons (Sikora, Begelman, & Rees 1994), photons from an accretion disk (Dermer & Schlickeiser 1993) or from a dusty torus (Sikora & Madejeski 2000). In the second class of models it is instead postulated that energy is mainly stored in hadrons which subsequently transfer it - through particle-particle or particle-photon interactions - to a population of leptons. The two components of the SED in these cases would be mostly synchrotron radiation from primary and secondary leptons or protons, respectively (Mannheim 1993, Protheroe 2001, Aharonian 2000).

Because of my personal bias against the second class of models (i.e. doubts concerning the radiative efficiency, variability timescales, spectral characteristics and relative variability of the two spectral components) hereafter I will focus on the leptonic (SSC and EC) scenario (see also Sikora & Madejski 2001).

2 A Blazar Sequence

Giommi & Padovani (1994) first pointed out that the different spectral properties/flavour of blazars (more specifically BL Lacs selected in different bands) could be accounted for by the different position of the low energy peak in the SED. Following their suggestion and independently of any model, Fossati et al. (1998) considered the largest complete blazar samples known at that time and divided the sources according to their radio luminosity (thought to trace well the bolometric power). This work revealed trends among the SED properties of the different classes of blazars: both the energy of the spectral peaks and the relative luminosity of the low vs high energy components – i.e. the relative importance of the two physical processes which produce them – systematically decrease with increasing source power (see Fig. 1). This trend in power also corresponds to a spectral sequence from the high frequency peaked (TeV candidates) BL Lacs (HBL) to the low frequency peaked (LBL) BL Lacs and to FSRQ (Fossati et

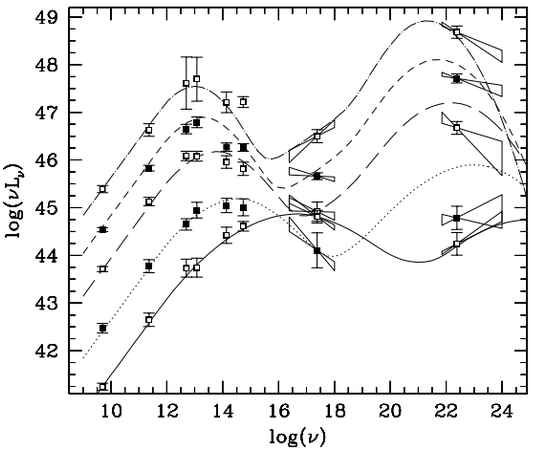

Fig. 1. The blazar sequence. The SEDs represent the average fluxes for complete samples of blazars where objects are 'divided' only according to their radio power. From Fossati et al. (1998).

al. 1998). Note that this behaviour might be somewhat surprising. Since the observed luminosity in these objects is enhanced by beaming by orders of magnitude, changes in viewing angle and/or bulk Lorentz factor would dramatically change the observed power. However the observed (bolometric) flux is $\propto \delta^4$ while the peak frequencies is $\propto \delta$, i.e. in direction opposite to the trend described above [1]. This might then indicate that the spread in angles and Lorentz factors are rather limited or that the velocity field of the emitting plasma spans a range in angles (i.e. it is not parallel to the jet axis) and blazars represent sources observed within such jet opening angle (Ghisellini 2001).

Let us examine in some more detail the properties of blazars along such phenomenological sequence.

Powerful blazars. These objects are among the most powerful sources known, with an isotropic power which exceeds 10^{49} erg s^{-1} (plausibly the radiation is largely boosted by beaming), and the second most powerful engines to produce bulk kinetic energy after Gamma–Ray Bursts (GRB).

These sources include FSRQ with relatively strong emission lines and several BL Lacs of the LBL–type. The synchrotron (low energy) peak is in the mm–far IR band, while the high energy one is located in the MeV band and largely dominates the power output (the SED of a typical source of this category, PKS 0836+710, is shown in Fig. 2). Powerful blazars are not very luminous in the optical, since their synchrotron peak is at lower frequencies, and thus they have the largest X–ray to optical flux ratio.

[1] δ is defined as the Doppler factor $\delta \equiv [\Gamma(1-\beta cos\theta)]^{-1}$, see Mastichiadis, this volume – *the eds.*

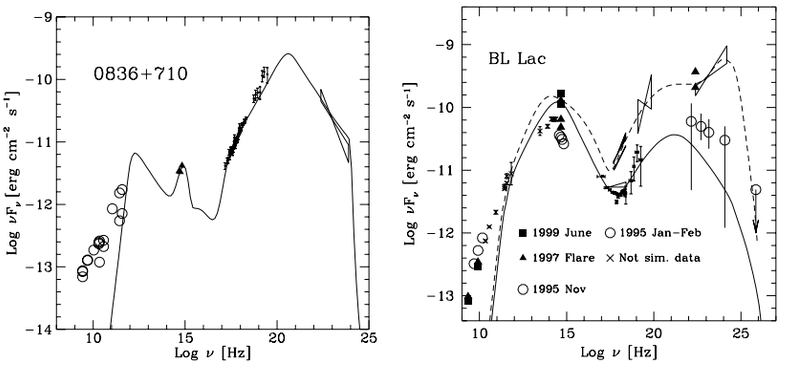

Fig. 2. Examples of SEDs of blazars. **Left:** SED of PKS 0836+710, a Flat Spectrum Quasar at $z = 2.17$ (from Tavecchio et al. 2000). Its power output is dominated by the high energy emission, as revealed by *Beppo*SAX and EGRET observations. **Right:** SED of BL Lac itself, a Low frequency peaked BL Lac object, in the flaring state of 1997 and in the *Beppo*SAX observation of 1999. During the latter, the optical flux was comparable to that during the 1997 flare, but the X–ray emission was much lower, revealing both the synchrotron and the inverse Compton components. From Tagliaferri et al. (2000).

Intermediate blazars. Fig. 2 also shows an example of intermediate blazar, BL Lac itself. Most of these objects are classified as LBL blazars and their X–ray spectrum often reveals both a steep power law (identified as the high energy tail of the synchrotron emission) and a very hard power law at higher energies (plausibly the emerging inverse Compton component). During the *Beppo*SAX observations of BL Lac, the source showed very rapid variability (timescales of 20 min) in the soft energy band, absent at higher energies. This behavior might be typical in these sources and could be due to different emitting zones or to different cooling times of the radiating electron (the most energetic ones are producing the synchrotron tail while the hard X–rays are emitted by electrons with much less energy and thus longer cooling times).

Weak blazars. These are the least powerful blazars, with the synchrotron peak located above the optical band, and sometimes reaching the hard X–ray band, as in Mkn 501 whose SED is shown in Fig. 3. Furthermore, these sources are the best candidates to emit copiously in the TeV band – as in the case of Mkn 421, Mkn 501, PKS 2155–304 and 1ES 2344+514.

High energy (especially in the X–ray and TeV bands) observations are important because this radiation is produced by the most energetic electrons (with random Lorentz factors $\gamma > 10^6$), which can then reveal properties of the acceleration mechanism at its extreme. Besides giving information on the emission and acceleration mechanism, TeV–band data can also allow the measurement of the amount of absorption (through the γ–$\gamma \rightarrow e^{\pm}$ process) both local to the

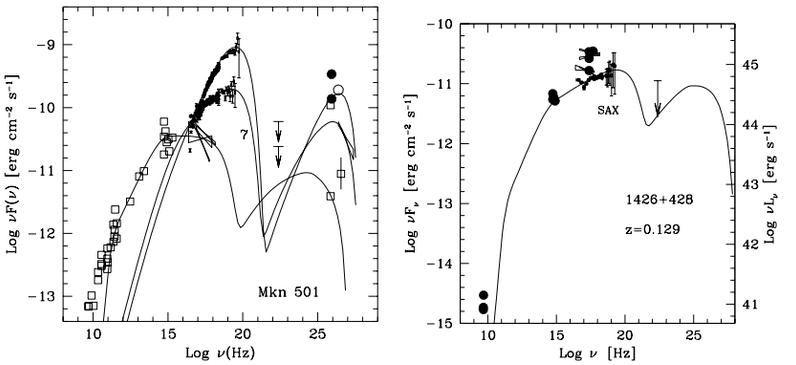

Fig. 3. Left: SED of Mkn 501, a High frequency peaked BL Lac object, during a quiescent state and the flaring state of 1997 (adapted from Pian et al. 1998). **Right:** SED of the blazar 1428+4217 at $z = 4.72$, showing a high energy component dominating the luminosity output during recent *Beppo*SAX observations. The line represents a model taking into account synchrotron and inverse Compton emission from a single homogeneous region. From Fabian et al. (2001b).

source and the one due to the infrared background radiation (see e.g. Stecker & De Jager 1997)

One can ask if even weaker sources can exist, with the synchrotron peak in the MeV range. According to the proposed blazar sequence, these should be low power objects with weak emission in all bands but the MeV one (i.e. radio–weak and dominated by the light of the galaxy in the optical). The SSC emission in the $>$ TeV band could be also faint because of Klein–Nishina effects. As suggested by Ghisellini (2001) therefore there may be many objects with overall low power emission, undetectable in 'inactive' elliptical galaxies except in the MeV band. Future observations in such band might therefore discover these "synchrotron MeV" BL Lacs.

Recent work has led to the construction of blazars samples both larger than and selected with criteria different from the 'classical' radio and X–ray samples. Interestingly such new samples (e.g. DXRBS, Sedentary, BLEIS, EMSS quasars, REX, RGB see e.g. Padovani & Urry 2001 and reference therein) have revealed a significant number of sources with intermediate properties (e.g. peak frequencies), but – so far – in global agreement with the above sequence.

2.1 A Physical Interpretation

Phenomenological evidence has thus lead to the proposal of a defined sequence in the blazar properties. But what is the physical origin of such a trend ?

Clearly the interpretation depends on the assumptions of the physical model reproducing the observed emission. In the leptonic scenario the blazar sequence corresponds to a decrease in the energy of electrons emitting at the SED peaks with increasing energy density of the soft photons seeds for Compton scattering

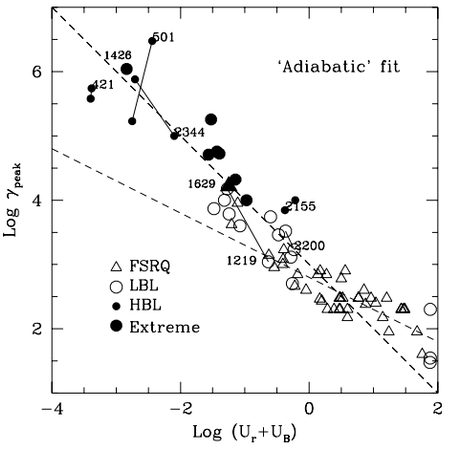

Fig. 4. The random Lorentz factors of the electrons emitting most of the blazar radiation, $\gamma_{\rm peak}$, as a function of the comoving (magnetic and radiative) energy density U. Note the two branches of the $\gamma_{\rm peak}$ vs U relation. At high U, $\gamma_{\rm peak}$ corresponds to the less energetic electrons continuously injected throughout the source, as the shape of the steady particle distribution appears to be dominated by the radiative cooling. At lower U (corresponding to extreme HBL sources) adiabatic losses become important giving rise to the $\gamma_{\rm peak} \propto U^{-1}$ branch. From Ghisellini & Celotti, in prep.

and source power (Ghisellini et al. 1998; see also Georganopoulos, Kirk & Mastichiadis 2001). More specifically, Ghisellini et al. (1998) applied to all blazars detected by EGRET with known redshift and γ–ray spectrum, a one–zone homogeneous synchrotron and inverse Compton model (considering also photons produced externally to the jet as possible seed photons for the inverse Compton process). This led to an estimate of the intrinsic source parameters namely the radiation energy density, the magnetic field, the beaming factor, the energy of the electrons emitting at peaks of the SED (represented by their random Lorentz factor $\gamma_{\rm peak}$) and the source size. They found that the best trend – underlying the spectral sequence – is a remarkable correlation between $\gamma_{\rm peak}$ and the total comoving energy density U (which includes magnetic and radiative energy densities). Such a trend has been later extended by considering more extreme sources (Ghisellini & Celotti, in prep.), i.e. extreme HBL blazars recently observed by *Beppo*SAX (although they lack information on the higher energy peak as not detected by either EGRET or by Cherenkov telescopes).

The overall behaviour $\gamma_{\rm peak}$ vs U is shown in Fig. 4. The correlation seems to be reasonably well approximated by two power laws. At higher U the slope is roughly $\gamma_{\rm peak} \propto U^{-0.6}$. This suggests that radiative cooling is playing a crucial role in determining this behaviour, as the radiative cooling rate is $\dot{\gamma} \propto U\gamma^2$. The found slope would thus imply that the cooling rate at $\gamma_{\rm peak}$ is almost the same in all sources.

The adopted model assumes that there is a continuous injection of relativistic particles throughout the source at a rate $\propto \gamma^{-s}$ (with $s > 2$) above some Lorentz

factor $\gamma_{\rm min}$. The resulting steady particle distribution can then be described by a broken power law with a peak at $\gamma_{\rm peak} = \gamma_{\rm min}$, as radiative cooling dominates at all energies.

However, at lower U – corresponding to the extreme HBL objects – the correlation is not compatible with $\gamma_{\rm peak} \propto U^{-0.6}$, but more similar to $\gamma_{\rm peak} \propto U^{-1}$ [2]. What happens is that as the radiative cooling timescales in these low power sources is longer, adiabatic losses become important even for electrons at $\gamma > \gamma_{\rm min}$. In this case the peak of the emission is located at energies for which the radiative and adiabatic cooling rates are equal, giving raise to a $\gamma_{\rm peak} \propto U^{-1}$ relation (Ghisellini & Celotti in prep).

3 Jet Power

The reasonably coherent picture of the blazar zoology described above has interesting consequences on the key issues of the origin and functioning of relativistic jets.

Due to beaming, it is not simple to even estimate the amount of luminosity intrinsically emitted by blazars and radio–galaxies and from it the power carried by jets. However, powerful radiogalaxies and quasars have extended radio structures containing huge energies, even $\sim 10^{59-60}$ erg – as calculated on the basis of equipartition arguments and simplifying assumptions. By dividing this energy by the estimated lifetimes of these structures it has been possible to derive the average power which had to be supplied through the jets (Rawlings & Saunders 1991). It appears that the limits on the power carried by jets on the extended scales (typically at 100–1000 kpc from the core) can be roughly comparable or even exceed the luminosity produced during the phenomenon of accretion. Thus the creation of jets has to be a relevant process in the functioning of the central engine.

Blazars show the power radiatively dissipated by jets on the smallest observable scales ($<$ pc): this has to be only a small fraction of the overall power transported by jets, for it to be mostly supplied to the extended structures. Nevertheless, the spectral properties of blazars can be used to significantly constrain - on the corresponding physical scales - both the powers involved, the energy carrier, i.e. the content of matter in the jets, and the related characteristics of the particle distribution(s).

In fact let us first consider the various forms of power associated to the blazar phenomenon (Celotti & Ghisellini, in prep).

A lower limit on the jet power can be estimated by simply dividing the observed luminosity by the square of the Lorentz factor, i.e. calculating the power effectively dissipated through radiation (assuming that the emission process is more or less continuous, with no accumulation and rapid release of random energy). The inferred values are of order of $L_{\rm rad} \sim 10^{47} L_{\rm obs,49}/\Gamma_1^2$ erg s^{-1}. Furthermore, the kinetic power transported by the emitting jet (i.e. at distances

[2] Note that more uncertainties affect the determination of the intrinsic parameters as no information on the bulk of Compton emission is available.

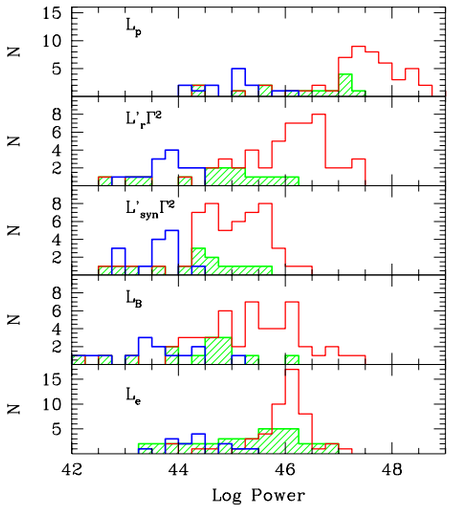

Fig. 5. Histograms of the distribution of powers (in erg s^{-1}) of the blazars considered in Celotti & Ghisellini, in prep: $L_{\rm p}$ is the power carried by protons assuming one proton per emitting electron; $L'_{\rm r}\Gamma^2$ is the power radiated by the jet and $L'_{\rm syn}\Gamma^2$ the power in synchrotron radiation only; $L_{\rm B}$ is the Poynting flux; $L_{\rm e}$ is the kinetic power carried by the emitting electrons. The gray histograms and shaded areas correspond to FSRQs and BL Lacs, respectively. The thick line histograms correspond to extreme BL Lacs (not detected by EGRET) for which we have assumed that the SSC emission has about the same power as the synchrotron one.

of 0.01-0.1 pc) can be computed by considering the density of particles required to account for the observed luminosity, by applying once again a simple one–zone homogeneous synchrotron plus inverse Compton model to the observed SED.

In Fig. 5, the distributions of $L_{\rm rad}$ for a sample of EGRET detected blazars is compared to the distributions in: bulk kinetic power in protons $L_{\rm p}$ (assuming one proton for each emitting electron), in emitting particles ($L_{\rm e}$) and the power as Poynting flux ($L_{\rm B}$). As can be seen, on average the dissipated power ($L_{\rm rad}$) is larger than those carried by emitting particles (electrons or electron–positron pairs) and magnetic field only. This therefore requires an energetically important proton component as the major energy carrier. The $L_{\rm p}$ distribution shown in the figure, which represents the values corresponding to a negligible amount of emitting pairs, is on average a factor 10 larger then the dissipated power. This in turn strongly limits the contribution of any electron–positron population.

The powers represented in the histograms are also based on the (strong) assumption that the emitting particle distribution extends down to $\gamma_{\rm min} \sim 1$, as this parameter determines the total particle number (but see below). However, $L_{\rm rad}$ does not depend on it, and thus in order for the kinetic power to exceed the dissipated one, $\gamma_{\rm min}$ cannot be larger than a few.

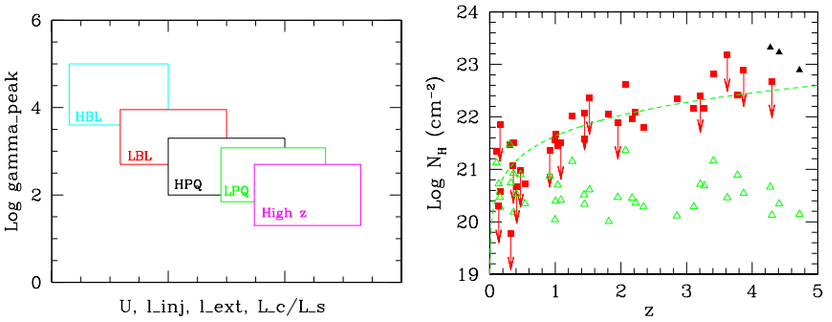

Fig. 6. Left: Schematic diagram representing the trend of decreasing energy of the blazar SED peaks with increasing energy density and power of the source, from low energy peaked, high energy peaked BL Lacs to flat spectrum quasars and high z objects. **Right:** Flattening of the soft X–ray spectrum (parameterized by an equivalent hydrogen column density N_H) vs redshift. The three triangles represent high z blazars. The other data are from Reeves & Turner (2000). The empty symbols indicate the Galactic column. Figures adapted from Ghisellini et al. (1998), Fabian et al. (2001a).

3.1 High Redshift Blazars

A few blazars at $z > 4$ have been recently identified and studied in the X–ray band (e.g. Fabian et al. 1997). Their global SED appear to be rather similar (considering the poorer spectral sampling) to those of nearby objects of similar power. However these high z sources present somewhat more extreme (model) properties, in terms of the required seed photon energy density and (low) $\gamma_{\rm peak}$. These objects thus naturally and consistently extend the trends described above to even higher blazar powers (see the schematic diagram in Fig. 6). Note that the location of the low energy peak makes the brightest sources ideal candidates to be detected by the future Planck satellite mission .

A further trend which appears to be related to the source power and/or redshift is a systematic increase of flattening in the soft X–rays compared to the extrapolation from the higher energies (typically 2-10 keV) power–law, as already noted by several authors for nearer objects (Elvis et al. 1998, Fiore et al. 1998). In Fig. 6 the flattening is represented by an increase in the amount of column density N_H of a (local) absorber with respect to z. The origin of such behavior is not clear. It could be accounted for by a change in the shape of the emitting particle distribution or the scattered soft photon field as well as by an effective increase in the column density of absorbing gas on nuclear or larger scales (Fabian et al. 2001a, 2001b).

On the basis of the (soft) X–ray observations of the most powerful blazars, in particular the high redshift ones, a further inference on the elusive parameter $\gamma_{\rm min}$ can be made: in fact while $\gamma_{\rm min}$ cannot be significantly constrained at low synchrotron frequencies (because of self–absorption), any low energy cutoff in the particle distribution at $\gamma_{\rm min}$ larger than unity would be visible in the low energy part of the inverse Compton spectrum, i.e. in the X–rays band for the

brightest sources, as pointed out by e.g. Sikora & Madejski (2000). Interestingly, all the observational information obtained so far on nearby and distant blazars supports the view that $\gamma_{\rm min}$ is of the order of unity (e.g. Fabian et al. 2001a, 2001b; Sikora & Madejski 2000; Tavecchio et al. 2000).

4 Internal Shock Scenario

While the above discussions on the properties of jets have been based on average steady state models, it is clear that variability, crucial characteristic of the observed radiation from blazars, might be also a key feature related to the central engine. In order to explore such a possibility the so called internal shock scenario has been considered (Rees 1978, Sikora, Begelman & Rees 1994, and for recent developments in GRB scenarios see Rees & Mészáros 1994) in which the central engine works intermittently, producing shells of slightly different physical properties. Faster shells emitted later can catch up slower (earlier) ones, forming shocks and dissipating part of their bulk kinetic energy into radiation.

Note that in the case of blazars only a moderate efficiency in the conversion from bulk to radiative energy is required, in good agreement with the predictions of such scenario (and at odds with the case of GRB). More generally this model appears to be able to account for some basic characteristic of blazars:

- The radiative efficiency – as just mentioned – is of the right order.
- If the initial separation between shells is of order of a few Schwarzchild radii (R_i), typically the shell–shell collision takes place at $R_i \Gamma^2$, which for $\Gamma \sim 10$ just corresponds to the distance where the inverse Compton scattering on external (broad line region) photons is mostly efficient, the γ–$\gamma \rightarrow e^{\pm}$ process is not important, and yet the region is still sufficiently small to account for the rapid variability.
- The global scenario would involve a hierarchical structure in shell–shell collisions: shells can collide again and again also at distances where the dominant emission is at low frequencies (via synchrotron radiation), thus predicting a connection between optical and γ–ray flares with radio–mm ones (e.g. see Fig. 7)..

These qualitative properties have been indeed reproduced via numerical simulations by Spada et al. (2001), where a jet of average kinetic power of 10^{48} erg s^{-1} is constituted of shells injected every few hours, with bulk Lorentz factors randomly chosen between 10 and 30. The first collisions occur well within a Broad Line Region (assumed to be located at 5×10^{17} cm and reprocessing 10% of a disc luminosity $\sim 10^{46}$ erg s^{-1}). Particles emit by synchrotron, SSC and EC. In Fig. 7 some spectra, corresponding to one single shell–shell collision at different distances are shown. [3].

Given the promising results of this scenario, further work is in progress in order to examine the whole blazar sequence.

[3] The entire time dependent evolution can be seen in the form of a movie at the URL: `http://www.merate.mi.astro.it/~lazzati/3C279/index.html`

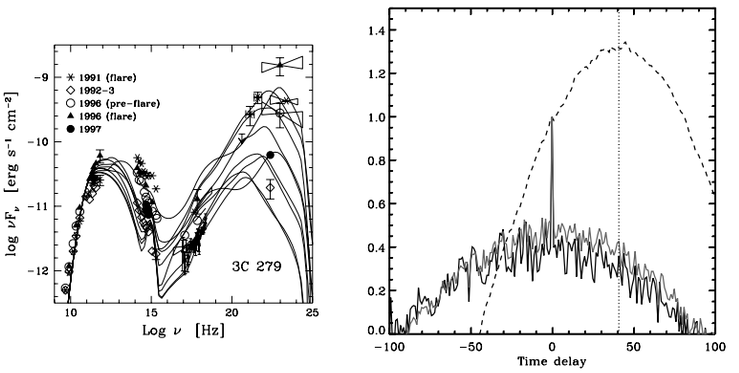

Fig. 7. Left: Cross correlation between the simulated γ–ray and X–ray, optical and mm light curves. Only for the latter there is a clear delay of about 40 days, due to the average distance of the regions producing γ–rays and mm radiation. From Spada et al. (2001). **Right:** SED of 3C 279 with, superimposed, different spectra resulting from simulations of internal shocks corresponding to collisions in different parts of the jet.

5 Extended (X-ray) Jets

Let us finally consider the recent discovery by *Chandra* of intense X–ray emission from large scale jets associated with quasars and radio galaxies (Chartas et al. 2000; Schwartz et al. 2000; Wilson et al. 2000). This in fact open the possibility of studying and significantly constraining the properties of jets (and ambient medium) on large scales.

In the specific case of the first detected objects, PKS 0637-752, such emission is produced at (de-projected) distances of $\sim$ Mpc from the central core. The relative intensity of X–ray to optical radiation implies that X–rays are likely not to be produced by the synchrotron process. A plausible alternative is that such emission is Compton scattered radiation of both/either the sub–pc core (blazar) spectral component and/or the cosmic microwave background (CMB) radiation. The energy density of the nuclear photon field dominates for a quasi–stationary (with respect to the observer) plasma component, while the CMB field would result enhanced and thus dominant for plasma in relativistic motion (see Fig. 8). It has been thus (re–)proposed the possibility that jets are constituted by different speed components, simplified as a central high velocity spine and a low velocity outer layer (see also Laing et al. 1999, Chiaberge et al. 2000). In turn, the low velocity/quasi-stationary emission would be seen also at large angles with respect to the jet axis (i.e. in radio galaxies), while the relativistically emitted radiation would be amplified for observers at small angles with the jet axis, i.e. quasars (see the resulting spectra in Fig. 8) (also Harris 2001; Wilson 2001, Celotti, Ghisellini & Chiaberge 2001). Such interpretation would then require that at least part of the jet moves at highly relativistic speeds up to the largest scales (thus also minimizing the required energy). Further *Chandra* observations will be crucial for testing such a possibility.

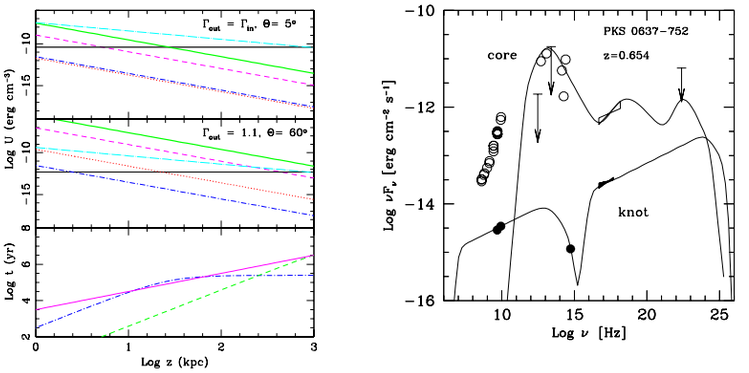

Fig. 8. **Left**: Energy densities as seen by a relativistically moving ($\Gamma \sim 10$, upper panel) and quasi-stationary (middle panel) plasma at various jet distances. Relevant energy densities are: from the inner blazar core (solid, oblique line), local synchrotron radiation (dot-dashed), CMB (solid, horizontal). The bottom panel shows the radiative (dash-dot for the spine and dashed for the layer) vs adiabatic (solid) losses. **Right**: SED of the nuclear (blazar) core and large scale together with the models for both components (see text). For more details see Celotti, Ghisellini & Chiaberge (2001).

6 Conclusions

The recent high energy observations of blazars (EGRET, TeV Cherenkov telescopes, *Beppo*SAX, HST) have revived very active research in the field as this conference has shown which will further benefit from the coming *Chandra* and XMM detailed images and spectra, and in the near future by the γ–ray satellites AGILE and GLAST.

Acknowledgments

I would like to thank the Organizers for setting up a very interesting meeting in a very pleasant atmosphere, for inviting me and for their patience during the production of these Proceedings. And it is a pleasure to thank Gabriele Ghisellini for years of fruitful collaboration. The Italian MURST is acknowledged for financial support.

References

1. F. A. Aharonian: New Astronomy, **5**, 377 (2000)
2. R. D. Blandford, M. J. Rees: in *Pittsburgh Conference on BL Lac Objects*, A.N. Wolfe ed., Pittsburgh University Press, 328 (1978)
3. A. Celotti, G. Ghisellini, M. Chiaberge: Mon. Not. R. Astr. Soc., 321, L1 (2001)
4. G. Chartas et al.: Astroph. J, **542**, 655

5. M. Chiaberge, A. Celotti, A. Capetti, G. Ghisellini: Astr. & Astroph., **358**, 104 (2000)
6. C. D. Dermer, R. Schlickeiser: Astroph. J., **416**, 458 (1993)
7. M. Elvis, F. Fiore, P. Giommi, P. Padovani: Astroph. J., **492**, 91 (1998)
8. A. C. Fabian, W. N. Brandt, R. G. McMahon, I. Hook: Mon. Not. R. Astr. Soc.,**291**, L5 (1997)
9. A. C. Fabian, A. Celotti, K. Iwasawa, C. L. Carilli, R. G. McMahon, W. N. Brandt, G. Ghisellini, I. Hook: Mon. Not. R. Astr. Soc., in press (2001a)
10. A. C. Fabian, A. Celotti, K. Iwasawa, G. Ghisellini: Mon. Not. R. Astr. Soc., in press (2001b)
11. F. Fiore, M. Elvis, P. Giommi, P. Padovani: Astroph. J., **492**, 79 (1998)
12. G. Fossati, L. Maraschi, A. Celotti, A. Comastri, G. Ghisellini: Mon. Not. R. Astr. Soc., **299**, 433 (1998)
13. M. Georganopoulos, J. Kirk, A. Mastichiadis: in *Blazar Demographics and Physics*, Padovani P., Urry M. eds., PASP, in press (2001)
14. G. Ghisellini: in *X-ray Astronomy 2000*, Giacconi R., Stella L., Serio S., eds., ASP Conference series, in press (2001)
15. G. Ghisellini, P. Madau: Mon. Not. R. Astr. Soc., **280**, 67 (1996)
16. G. Ghisellini, A. Celotti, G. Fossati, L. Maraschi, A. Comastri: Mon. Not. R. Astr. Soc., **301**, 451 (1998)
17. P. Giommi, P. Padovani: Astroph. J., **444**, 567 (1994)
18. D. Harris: in *Particle and Fields in Radio Galaxies*, Laing R., Blundell K. eds., ASP, in press (2001)
19. R. C. Hartman et al.: Astroph. J. Suppl., **123**, 79 (1999)
20. R. A. Laing, P. Parma, H. R. de Ruiter, R. Fanti: Mon. Not. R. Astr. Soc., **306**, 513 (1999)
21. K. Mannheim: Astron. & Astroph., **269**, 67–76 (1993)
22. P. Padovani, M. Urry: in *Blazar Demographics and Physics*, Padovani P., Urry M. eds., PASP, in press (2001)
23. D. Petry et al.: Astron. & Astroph., **311**, L13–16 (1996)
24. E. Pian et al.: Astroph. J., **491**, L17 (1998)
25. R. Protheroe: in *Particle and Fields in Radio Galaxies*, Laing R., Blundell K. eds., ASP, in press (2001)
26. S. Rawlings, R. D. E. Saunders: Nature, **349**, 138 (1991)
27. M. J. Rees: Mon. Not. R. Astr. Soc., **184**, P61 (1978)
28. M. J. Rees, P. Mészáros: Astroph. J., **430**, L93 (1994)
29. J. N. Reeves, M. J. L. Turner: Mon. Not. R. Astr. Soc., **316**, 234–248 (2000)
30. D. A. Schwartz et al.: Astroph. J., **540**, L69 (2000)
31. M. Sikora, M. C. Begelman, M. J. Rees: Astroph. J., **421**, 153 (1994)
32. M. Sikora, G. Madejski: Astroph. J., **534**, 109 (2000)
33. M. Sikora, G. Madejski: in *High Energy Gamma-Ray Astronomy*, F. Aharonian, H. Voelk eds., AIP, in press (2001)
34. M. Spada, G. Ghisellini, D. Lazzati, A. Celotti: Mon. Not. R. Astr. Soc., submitted (2001)
35. F. W. Stecker, O. C. De Jager: Astroph. J., **476**, 712 (1997)
36. G. Tagliaferri et al.: Astron. & Astroph., **354**, 431 (2000)
37. F. Tavecchio et al.: Astroph. J., **543**, 535 (2000)
38. T. C. Weekes et al.: Astron. & Astroph. Suppl., **120**, 603 (1996)
39. A. S. Wilson: in *Particle and Fields in Radio Galaxies*, Laing R., Blundell K. eds., ASP, in press (2001)
40. A. S. Wilson, A. J. Young, P. L. Shopbell: Astroph. J., **544**, L27 (2001)

Relativistic Outflows from X-ray Binaries ('Microquasars')

Rob Fender

Astronomical Institute 'Anton Pannekoek' and Center for High Energy Astrophysics, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

Abstract. In this review I summarise the observational connections between accretion and relativistic outflows – jets – in X-ray binaries. I argue that jets are likely to be a fairly ubiquitous property of X-ray binaries as a whole, an assetion which can be tested by further observations of the Atoll-type X-ray binaries. I discuss broad patterns that are emerging from these observational studies, such as a correlation between 'hard' X-ray states and the presence of radio emission, and the related anti-correlation between jet strength and mass-accretion rate as inferred from X-ray studies alone. I briefly discuss possible future directions for research and compare X-ray binary jets to those from Active Galactic Nuclei and Gamma Ray Bursts.

1 History and Introduction

There has been a great deal of renewed interest in the past half a decade or so in the phenomena of relativistic outflows, or 'jets' from binary systems in our own galaxy. These are often referred to as 'microquasars' because of the apparent similarities with the Quasars, or with Active Galactic Nuclei (AGN) in general. The particular type of stellar binary systems in which these jets seem to orginate are the X-ray binaries (XRBs), so-called because they are powerful sources of X-ray radiation. In XRBs a more-or-less 'normal' star loses matter to a compact collapsed companion, either a neutron star or a black hole; it is generally accepted that the accretion of material by the compact star, a process far more energetically efficient than nuclear fusion, is the source of the enormous power output of these systems (which can exceed in some cases 10^{38} erg s^{-1}, or the Eddington luminosity for a one solar mass object). Since the power source of AGN is similarly believed to be accretion of matter by a collapsed object, in this case a supermassive (10^6–10^9 $M_\odot$) black hole, the term 'microquasar' is more than simply an indicator of similar morphologies (ie. accretion, jet) but maybe also of similar physics. Therefore understanding such sources is important not only in the context of accretion physics and the evolution of 'local' systems, but maybe also for our broader understanding of the physics and evolution of the powerhouses of the universes, the AGN.

Fig 1 is a schematic of a generic X-ray binary, indicating the (probable) sites from which emission at different energies originates. Note that the observable spectral extent of such systems can be very broad – the classical black hole candidate (BHC) XRB, Cyg X-1, is a well-detected source from ≤ 1 GHz in the radio band to ≥ 1 MeV in γ-rays, a range of 10^{12} in photon energy. Most

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 101–122, 2002.

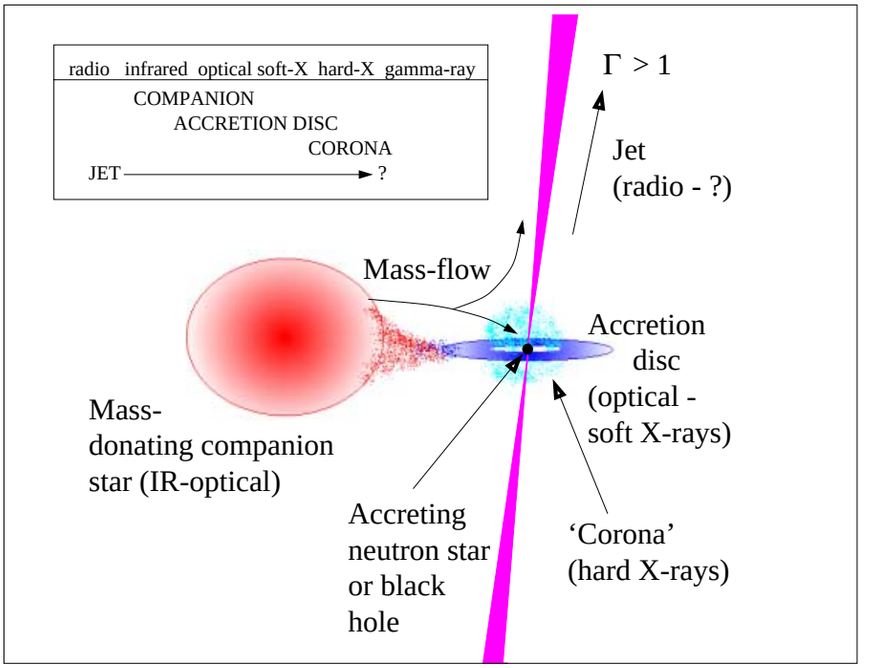

Fig. 1. A schematic of the generally accepted structure of a 'typical' X-ray binary system, indicating the locations of the sites believed to correspond to observed emission at different wavelengths.

schematics, certainly until a few years ago, would not have included the jet, and one of the goals of this paper will be to discuss exactly how ubiquitous is this feature in XRBs. It is interesting to note that the jet, when present, is by far the largest structure directly associated with the XRB in general, and accretion process in particular, and the only one of the structures indicated in Fig 1 which has actually been directly observed (in radio images – e.g. Fig 2).

Jets from XRBs as a phenomenon were first discovered from SS 433 (Spencer 1979; Hjellming & Johnston 1981a,b), a highly unusual system in many ways. The source displays optical (and infrared and X-ray) emission lines which show periodic Doppler shifts indicating a precessing bipolar outflow with velocity $v = 0.26c$; the radio jets appear to precess as predicted from the optical lines. The (apparently[1]) rather well-defined and only mildly relativistic velocity (bulk Lorentz factor $\Gamma = (1 - v^2/c^2)^{-1/2} = 1.04$) are unique amongst 'relativistic' jet sources. Importantly, SS 433 is the only system, XRB or AGN, for which atomic emission lines have been associated with the outflow, thereby establishing a baryonic content (more of this later). The significant and variable linear

[1] scepticism is healthy

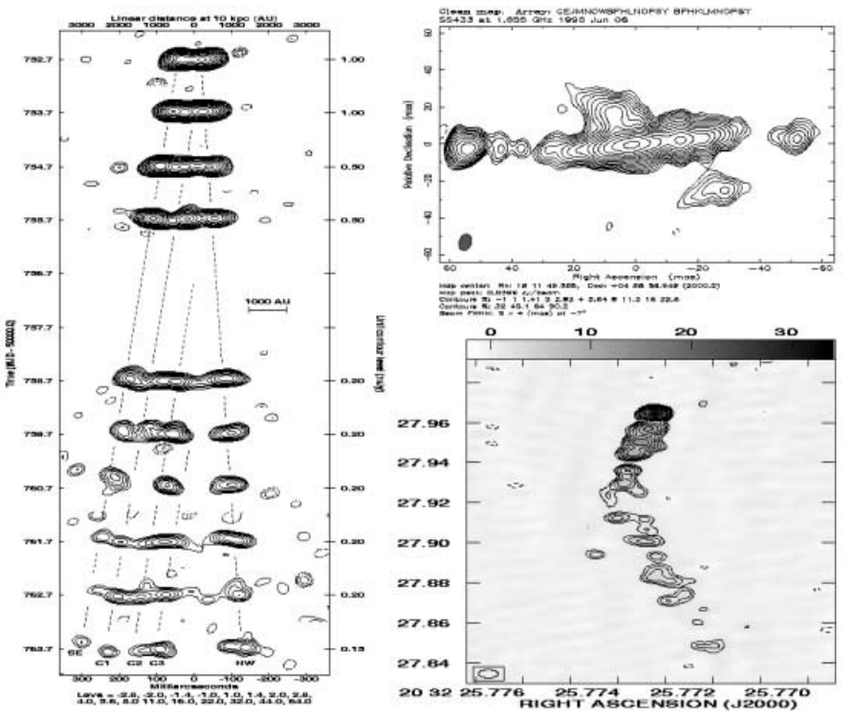

Fig. 2. Recent radio observations of three famous XRB jet sources. Left: A sequence of images of (apparent superluminal) ejections from GRS 1915+105 observed with MERLIN (Fender et al. 1999a). Top right: A recent combined EVN/VLBA image of SS 433 (Paragi et al. 2001). Lower right: a VLBA image of a one-sided curved jet from Cyg X-3 following a major radio flare (Mioduszewski et al. 2001).

polarisation of the jets confirmed the synchrotron interpretation for the origin of the radio emission.

Over the subsequent 15 years, a handful of other XRBs (e.g. Cyg X-3, Cir X-1, 1E 1740.7-2942, GRS 1758-258), were identified as being associated with radio jets. In the case of Cyg X-3, an apparent velocity of $\sim 0.3c$ was measured (Geldzahler et al. 1983). Perhaps all XRB jets would turn out to have a velocity of $\sim 0.3c$? This picture was comprehensively refuted in 1994 when Mirabel & Rodríguez (1994) discovered apparent superluminal motions in multiple ejections from the XRB GRS 1915+105. While distance-dependent, interpretations of the intrinsic velocity of the ejecta suggested $v \geq 0.9c$, significantly relativistic ($\Gamma \geq 2$). Clearly XRBs could eject material at extremely high velocities, comparable to those observed in AGN (where the phenomenon of apparent superluminal motion is relatively commonly observed and relatively easily explained as a geometric effect – e.g. Rees 1966; Zensus & Pearson 1987; Gomez et al. 2000). Shortly after the observations of GRS 1915+105, a second 'superluminal' XRB, GRO J1655-50, was discovered (Tingay et al. 1995; Hjellming & Rupen 1995). Since GRO

J1655-40 was demonstrated to be a strong BHC (Bailyn et al. 1995 and several subsequent papers), it was widely concluded that GRS 1915+105 was a BHC, something supported but never confirmed dynamically by further observations. Less certainly, it was asserted that the apparent dichotomy between the $\sim 0.3c$ sources (i.e. SS 433 and Cyg X-3) and the $\geq 0.9c$ sources (ie. GRS 1915+105 and GRO J1655-40) reflected the difference in escape speeds from the vicinity of neutron stars and black holes respectively (e.g. Livio 1999). However, at this stage it was still generally perceived that relativistic jets, as a property of X-ray binaries, were a rare phenomenon, a feature common only to a small group of 'unusual' systems. Recent radio images of three of the most famous jet sources are presented in Fig 2. It is worth reminding the reader that, as in the cases of AGN and GRBs, the outflows are relativistic in *two* senses – ie. they have relativistic ($1 < \Gamma < 100$) bulk velocities (ie. the proper motions we can resolve in radio images) and in addition are comprised of populations of relativistic particles ($1 < \gamma < 10000$) which, in spiralling around field lines in the magnetised plasma, produce the observed synchrotron emission.

It now seems likely that jets from XRBs are not so rare, and that for certain broad classes of X-ray binaries, maybe even the majority, the jet is as integral a part of the mass transfer process as the accretion disc. In this review I shall try to summarise the state of existing knowledge, and what appear to be fruitful avenues for further observational and theoretical study. The key question before we can advance to this stage is however, just how important are jets for the physical processes occurring in X-ray binaries? I hope to answer this in the next section by establishing their near-ubiquity.

2 The Near-Ubiquity of Jets from X-ray Binaries

In the following, I shall make the assumption that any detection of radio emission corresponds to evidence for jet production (in much the same way as detection of X-rays is taken as evidence for accretion processes). This is based upon the qualitative argument that whenever we have resolved radio emission it has had a jet-like appearance (except, perhaps, in the case of the unusual transient CI Cam in which a more isotropic radio nebula seems to have formed – Mioduszewski et al., in prep), and on the following more quantitative argument – for a maximum brightness temperature of $\sim 10^{12}$K for the synchrotron process, a flux density of ~ 5 mJy at 5 GHz (quite weak) corresponds to a physical size of $\sim 10^{12}$ cm at a distance of 5 kpc. This is an order of magnitude larger than the typical orbital separation ($\sim 10^{11}$ cm) of a low-mass X-ray binary, and so the simplest explanation is that such a large stucture is maintained by an outflow.

In terms of accretion and jet production, the most useful separation into classes of X-ray binaries is between neutron stars and black holes. I have attempted to do this both in the following sections, and also in table 1, in which approximate numbers of radio detections as a fraction of total known populations is indicated. Catalogues of XRBs with broad classifications can be found in van Paradijs (1995) and Liu, van Paradijs & van den Heuvel (2000,2001).

Table 1. Approximate numbers of radio detections (=jet production) in the different types of XRBs. Clearly detection of a large number of Atoll sources holds the key to unambiguously establishing (or not) the ubiquity of jets in XRBs.

Class	Fraction as radio sources
BHCs (persistent)	4 / 4
BHCs (transient)	$\sim$ 15/35
NS (Z)	6 / 6
NS (Atoll)	$\sim$ 5/100
NS (XRP)	$\sim$ 0/80

Note that while the fraction stated in table 1 for the BHC transients is $\sim$ 15/35, for those transients for which *any* radio observation was reported, the fraction is 15/16 (Fender & Kuulkers 2001) – how much this reflects non-publication of non-detections is unclear.

2.1 Neutron Stars

The three broad classes of accreting neutron star, and the relation between X-ray and radio properties, are summarised in Fig 3.

The Z Sources. These are the brightest persistent X-ray sources in the sky (the single brightest nonsolar X-ray source, Sco X-1, is the prototype of the group; Hasinger & van der Klis 1989). There are six in our galaxy, and possibly one in the LMC. The Z sources are thought to contain neutron stars with relatively low ($\leq 10^9$ G) dipole magnetic fields, accreting at or near the Eddington limit. All six galactic systems are variable but reliable radio sources, with a comparable radio luminosity (when on the 'horizontal branch' – Penninx 1989; Fender & Hendry 2000). The 'Z' refers to the pattern traced out in the X-ray colour-colour diagram (CD) in which three (possibly four) branches smoothly connect. Penninx et al. (1988) found that the radio emission was strongest on the 'Horizontal Branch' and weakest on the 'Flaring Branch' in the Z source GX 17+2, an apparent anti-correlation with accretion rate, $\dot{m}$ as deduced from X-ray observations alone. This relation between X-ray 'state' (as described by the branches of the Z) and radio emission seems to be a property common to all the Z sources (Hjellming & Han 1995 and references therein). Recently high-resolution radio observations have revealed unequivocal evidence for a variable jet-like structure associated with Sco X-1 (Bradshaw, Fomalont & Geldzahler 1999; Fomalont, Geldzahler & Bradshaw 2001a,b). The observations, of a variable core and moving, variable lobes, are interpreted as the impact of a highly relativistic beam on the ISM, producing the advance of radio 'hotspots' (Fomalont et al. 2001a,b). Given the similarities in their radio properties, and the resolved jet in Sco X-1, the simplest conclusion (ie. using Occam's razor) is that all Z sources produce relativistic jets. Furthermore, the 'unusual' system Cir X-1 has some Z-like properties (Shirey, Bradt & Levine 1999) and is a source of radio jets from arcsecond to arcmin angular scales (Stewart et al. 1993; Fender et al. 1998).

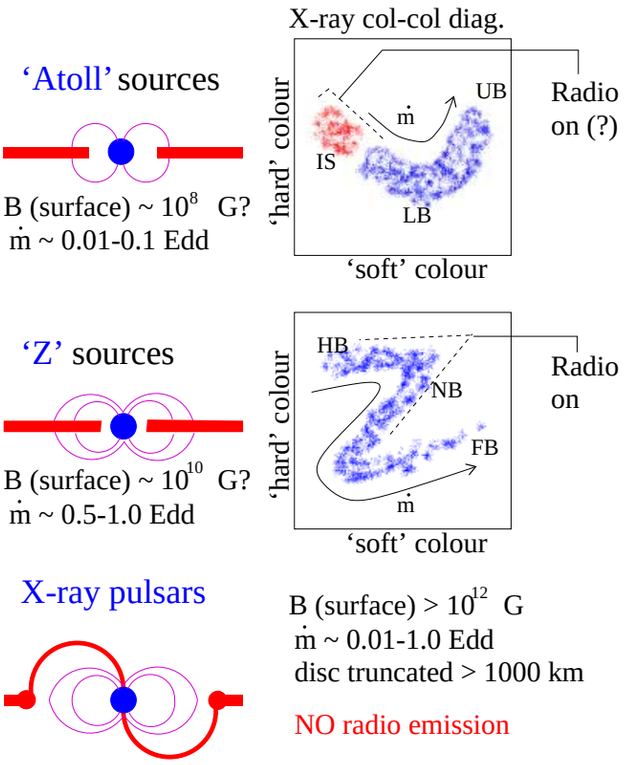

Fig. 3. A rough schematic of the three types of neutron star X-ray binary, a rough physical interpretation, and indication as to which states are associated with the presence of radio emission. For the Atoll sources, this association is currently speculative. Abbreviations are IS=Island State, LB=Lower Banana, UB=Upper Banana, HB=Horizontal Branch, NB=Normal Branch, FB=Flaring Branch.

The Atoll Sources. In the original classification of Hasinger & van der Klis (1989) and subsequent works, the Atoll sources were discussed as a separate small subgroup of bright low mass X-ray binaries with relatively low magnetic fields, accreting at lower rates than the Z sources (e.g. van der Klis 1995 for more details). Since then it seems likely that Atoll-like properties may be shared by the majority of low-field accreting neutron stars (van Paradijs, Ford, van der Klis, private communication) and so we shall adopt this viewpoint here. If this is the case then the Atoll sources, which will now include the groups of 'bursters', 'dippers' etc., are the largest class of catalogued X-ray binaries (see table 1). Little is known about the radio properties of the Atoll sources, except that, as a population, they are faint sources (typically < 1 mJy at cm wavelengths – Fender & Hendry 2000). The only Atoll source to be regularly and repeatedly detected at radio wavelengths was, until recently, the bright system GX 13+1 (Garcia et al. 1988). More recently other Atoll systems have been discovered

to have radio counterparts (e.g. 4U 1728-34/GX 354+0 – Martí et al. 1998), and sources with known transient radio counterparts have been discovered to be Atoll-like in nature (e.g. Aql X-1 – Reig et al. 2000). So while it is clear that as a population Atoll sources are not particularly radio-bright, it is also clear that they do produce detectable radio emission uncer certain conditions. This then implies that the majority of catalogued low-mass X-ray binaries are capable of producing a radio jet. However, until a jet is directly resolved from an Atoll source (a key future observation) this will remain unproven.

The X-ray Pulsars. These systems possess much stronger magnetic fields ($\geq 10^{11}$ G) than the Z or Atoll sources, which results in the disruption of the accretion flow at a distance of several thousand km from the neutron star surface (e.g. Bildsten et al. 1997 and references therein). As a population they are significantly fainter than even the Atoll sources, and no strong-field X-ray pulsar has ever been detected as a radio synchrotron source (Fender & Hendry 2000). Thus the strong possibility exists that such systems do not produce jet-like outflows, due to the extreme disruption of the accretion flow. Deep radio observations of some nearby X-ray pulsars would be useful to further constrain this.

NS Transients. Neutron star soft X-ray transients (Chen, Shrader & Livio 1997; Campana et al. 1998) can probably be classified as Atoll-like (e.g. the case of Aql X-1 – Reig et al. 2000). As with the BHC transients, there seems to be a discrete ejection of synchrotron emitting material associated with the sudden increase in luminosity at the start of the outburst. This manifests itself in a transient radio event which becomes optically thin within a few days (presumably due to decreasing self-absorption as the ejected component expands) and then fades away monotonically (Hjellming & Han 1995; Fender & Kuulkers 2001).

Several transients contain high field accreting X-ray pulsars (e.g. Bildsten et al. 1997) and, as with the more persistent sources of this type, none have ever been detected as radio synchrotron sources (Fender & Hendry 2000).

2.2 Black Hole Candidates

The description of the accretion state of the BHCs differs from that of the neutron stars as it is perceived that probably all black holes can, under the right conditions, achieve all states – ie. it is not currently perceived that there are different 'types' of black hole (the only obvious distinguishing characteristic would be the black hole spin). These 'states' are summarised in Fig 4. Note that while it was originally supposed that the states were a more or less one-dimensional function of mass accretion rate, which could itself be tracked via soft X-ray (disc) flux (see e.g. the pattern of behaviour in GRO J1655-40 – Mendez, Belloni & van der Klis 1998), it now seems clear that the picture is not so simple. One problem is that the same 'state' in terms of the X-ray spectral and timing properties can be reproduced at extemely different flux levels (e.g. Homan et al. 2001).

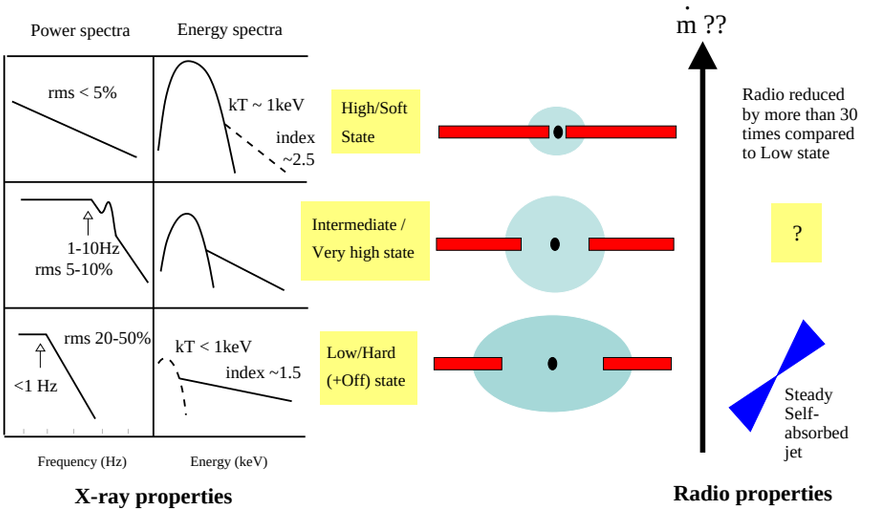

Fig. 4. BHC spectral states, as classified by their X-ray spectral and timing properties. A rough physical interpretation, based on the X-ray data alone, is indicated, as is the relation to radio emission. BHC transients typically, although not exclusively, transit from undetectable levels to the soft state in a short period of time. The relation of states to mass accretion rate, previously thought to be quite clear, is now less certain.

The two most distinct states are the the Low/Hard and High/Soft states, being the extremes of 'nonthermal' and 'thermal' spectra respectively (this is an oversimplification). There also exists a hybrid state, labelled the Very High or Intermediate state, which is less commonly observed than either the Low/Hard or High/Soft states, whose relation to radio emission, and hence presumably jet production, is unclear (but see Corbel et al. 2001).

The Low/Hard State. The Low/Hard X-ray state (historically called 'Low' because it is generally weaker than the High/Soft state in the soft X-ray band, and 'Hard' since it is dominated by a nonthermal power-law component which peaks at hard X-ray (≥ 50 keV) energies) is the state in which the four persistent BHCs in our galaxy spend most of their time (I consider these four to be Cyg X-1, GX 339-4, 1E1740.7-2942 and GRS 1758-258, although it should be noted that, at the time of writing, GX 339-4 has been at extremely low levels for over a year and consistently displays a larger amplitude of X-ray variability than the other systems).

In the early 1970s a transition from the High/Soft (possibly only 'Intermediate' – see discussion in Belloni et al. 1996) to Low/Hard X-ray states in Cyg X-1 was observed to be coincident with the appearance of a radio counterpart to this source (Tananbaum et al. 1972). It has since been established that while the source is in the Low/Hard state, which seems to be most of the time, it steadily emits a relatively low level (typically 5-15 mJy at cm wavelengths) of

radio emission (e.g. Brocksopp et al. 1999). The spectrum of the radio emission is remarkably flat and extends to at least the millimetre regime (Fender et al. 2000b). Furthermore, the radio emission is modulated at the 5.6-day orbital period of the system (Pooley, Fender & Brocksopp 1999). All of this evidence taken together suggests that the flat spectrum radio emission arises in a continuously-generated, partially self-absorbed compact jet from the system (with the orbital modulation possibly due to variable free-free absorption in the dense stellar wind of the OB-type mass-donor – Brocksopp 2000). Confirmation of this hypothesis appears to have recently been achieved with VLBA images of the system clearly resolving an asymmetric jet from a compact core (Stirling, Garrett & Spencer 1998; Stirling et al. 2001).

The other three persistent Low/Hard state systems also show flat radio spectra, and the two Galactic centre sources, 1E1740.7-2942 and GRS 1758-258, are associated with parsec-scale jet/lobe structures (the original motivation for the name 'microquasar'; Mirabel et al. 1992; Rodríguez, Mirabel & Martí 1992; Mirabel 1994). Furthermore, in both Cyg X-1 and GX 339-4 there is an approximately linear relation between the X-ray flux (dominated by the non-thermal power-law) and the radio emission (Brocksopp et al. 1999; Corbel et al. 2000) indicating a clear coupling between accretion (presumed to be reflected in the strength and spectrum of the X-ray emission) and outflow. While GX 339-4 is the one source for which the jet has probably not been reliably resolved, it is the one for which linear polarisation (at the level of a few %), has been measured (Corbel et al. 2000), supporting the synchrotron-emitting jet model.

While most BHC transients are observed to evolve rapidly (hours) from a 'quiescent' state to the High/Soft state (and are generally accompanied by an optically thin radio outburst – e.g. Fender & Kuulkers 2001), a few X-ray transients have been observed to spend an extended period in the Low/Hard state. A careful comparison of these Low/Hard state transients reveals that, following an initial optically thin radio event, they develop low-level, inverted-spectrum radio components. These were originally dubbed 'second stage' radio sources (e.g. Hjellming & Han 1995 and references therein). In Fender (2001) it is argued that these components are the same as the flat/inverted spectrum components observed from the persistent sources in the Low/Hard state, and furthermore that such spectral components, and therefore compact jets, are a general property of the Low/Hard state.

The High/Soft State. Early observations of Cyg X-1 (e.g. Tananbaum et al. 1972) suggested that the radio emission from the source was suppressed when in the High/Soft state (compared to the Low/Hard state). Despite the great accessibility of this system to the world's radio telescopes, a chance to test this hypothesis during a transition of the source to a soft X-ray state in 1996 was missed. However, observations near the end of the soft state support a scenario in which radio emission is stronger during the Low/Hard state (Zhang et al. 1997).

It was a year-long transition to the High/Soft state by GX 339-4 in 1998 in which the 'quenching' of the radio emission compared to the Low/Hard state was definitively established (Fender et al. 1999; Corbel et al. 2000). Radio monitoring of the source revealed the cm wavelength radio emission to have dropped by a factor of ≥ 25 during the High/Soft state, and to return to its previous levels once the source resumed the Low/Hard state (Fig 5).

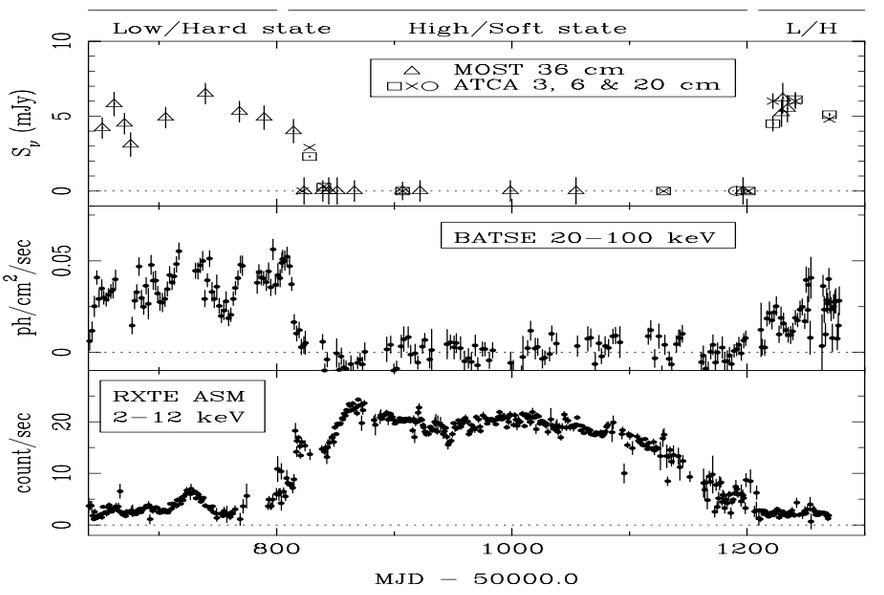

Fig. 5. Simultaneous 'quenching' of the radio (top panel) and hard X-ray emission during a year-long high/soft state in the black hole candidate GX 339-4 (from Fender et al. 1999; see also Corbel et al. 2000).

The Very High/Intermediate State. Little is clearly understood about the radio emission during the (comparitively rare) Very High/Intermediate state of BHCs, in which both thermal (disc) and nonthermal (power-law) spectral components can be present. Is it the presence of the nonthermal component, or the absence of the thermal component, which is necessary for the production of radio emission? Recent observations (Corbel et al. 2001) suggest the latter, but further observations of this state are required.

BHC Transients. BHC X-ray transients (e.g. Chen et al. 1997; Charles 1998) are generally associated with radio outbursts (e.g. Hjellming & Han 1995; Fender & Kuulkers 2001), which have been resolved on a small number of occasions into discrete ejections, sometimes multiple, of radio emitting components (e.g. GRO J1655-40 – Tingay et al. 1995; Hjellming & Rupen 1995).

In most cases the transients seem to transit from 'quiescence' (which may be some very low-level version of the Low/Hard state described above) to the High/Soft state in the space of a few days or less. However, in some rare cases (e.g. Fender 2001; Brocksopp et al. 2001) a transient will 'only' make it to the Low/Hard state. Whichever 'branch' the transient takes, it seems that the intial rise is generally associated with a discrete ejection event (Fender & Kuulkers 2001), sometimes multiple events (e.g. Kuulkers et al. 1999). Subsequently, if the source 'achieves' the High/Soft state, there appears to be no re-emergence of core radio emission (and so we can assume the radio jet is 'switched off' and any radio emission we observe is physically decoupled from the ongoing accretion process); if it instead finds itself in the Low/Hard X-ray state a flat or inverted-spectrum component emerges (Fender 2001).

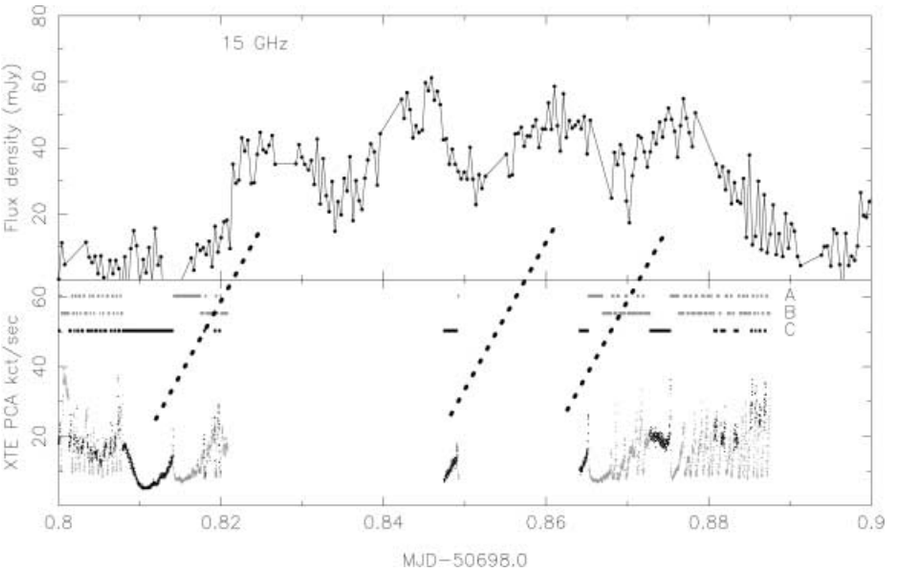

Fig. 6. The one-to-one correspondence, in GRS 1915+105, of brief (minutes) transitions into state 'C', roughly analogous to the Low/Hard state in more traditional BHCs, with the formation of discrete radio oscillation events. Four radio events are observed to be associated with a dips into the hard state – gaps in the X-ray light curve are due to Earth occultations, and there was almost certainly a hard dip associated with the second radio event as well. The delay is probably due to the time required for self-absorption of the radio emission to decrease (as the ejecta expand). From Klein-Wolt et al. (2001).

GRS 1915+105. While one of the aims of this review is to establish that jets are not a bizarre property of some small subset of XRBs, but rather a more ubiquitous feature, one source, GRS 1915+105, still deserves a mention on its own.

The source displays a remarkable and unique range of X-ray behaviour which can however be broken down into transitions between three broadly-defined 'states' (Belloni et al. 2000). It was also the first system for which we had direct evidence of highly relativistic flows (Mirabel & Rodríguez 1994) and displays an extraordinary variety of radio behaviour (e.g. Pooley & Fender 1997), most of which can (presumably) be associated with the formation of synchrotron-emitting jets.

One of the three 'states' into which any X-ray light curve of GRS 1915+105 can be deconstructed (state 'C') is broadly analogous to the Low/Hard state of traditional BHCs, being dominated by a power-law component in the X-ray band. From a comparison of many hours of overlapping X-ray and radio observations, we are confident that radio oscillation events are directly, and only, associated with these hard states in this source (Fig 6; Klein-Wolt et al. 2001), although there are alternative opinions expressed in the literature (e.g. Naik & Rao 2000).

GRS 1915+105 shows other extraordinary properties, as if attempting to provide all the observational data we need to understand the 'disc–jet' coupling on its own. For example it was the first source for which there was unequivocal evidence for infrared synchrotron emission (Fender et al. 1997; Eikenberry et al. 1998, 2000; Mirabel et al. 1998; Fender & Pooley 1998; 2000) and is the clearest example of a flat–spectrum core being resolved into a quasi-continuous jet (Dhawan, Mirabel & Rodríguez 2000; Feretti et al. 2001).

3 Connections

Some broad patterns are now beginning to emerge from these studies; these patterns are clues to generic properties of jets, their coupling to the accretion process and so on. In no particular order, these include:

- A broad correlation between hard X-ray states and radio emission, in particular in BHCs. Meier (2001) takes such observations as direct evidence for the MHD formation of jets in geometrically thick accretion flows. These hard X-ray states are generally interpreted as arising via inverse Comptonisation (Poutanen 1998 and references therein) in a 'corona' and/or an advection-dominated accretion flow (ADAF; e.g. Esin, McClintock & Narayan 1997). It is interesting to note that, to my knowledge, in every case where an ADAF has been invoked to explain the optical–X-ray spectrum in XRBs, radio emission is present (and yet is not fit by ADAF models).
- A related point is that in all types of X-ray binary for which we have a clear picture, the strength of the jet appears to be *anticorrelated* with the mass accretion rate as inferred from X-ray spectral and timing studies alone (e.g. Figs 2,3; Belloni, Migliari & Fender 2000). Jets may well turn out to be an important factor in the state transitions associated with BHCs, Z sources and probably also Atolls.
- The jet, however it is formed, whether via MHD or some other 'black box' really seems to carry a lot of the accretion power – in the case of the hard state of BHCs it seems inescapable that the jet requires at least 10% of the accretion

energy budget – since current models often attempt to fit observations to a rather higher degree of accuracy than this its effect can presumably not be ignored.

4 Forwards

In Fig 7 I attempt to briefly summarise the state of play of research into X-ray binary jets. A key point is the ubiquity of jets; it is hoped that I have by now convinced the reader that it is at least *possible* if not *likely* that jets are important for the majority of X-ray binary systems. Taking this as established, I have outlined the areas in which important research can be done via either

- An empirical, 'energetics' approach, in which the exact way in which the jet extracts energy and angular momentum from the accretion flow is not key,

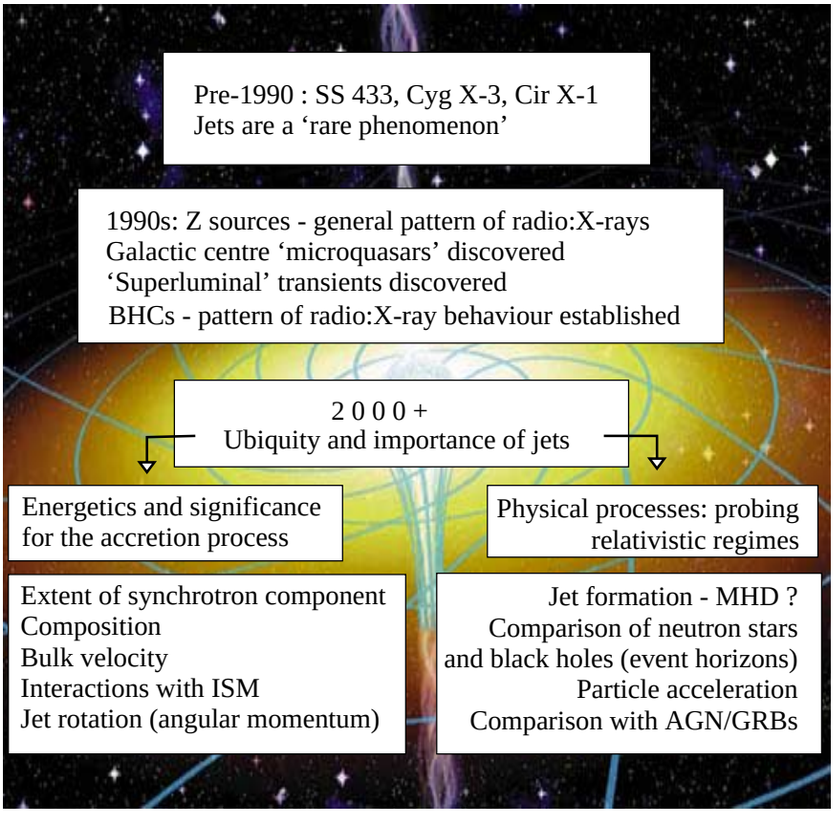

Fig. 7. A schematic indication of where we are now with research into jets from X-ray binaries, and some possible future directions. Central to this scheme is an understanding of how ubiquitous is the jet phenomena for X-ray binaries.

but rather estimating *how much* of these quantities are associated with the jet, is.

- A 'physics' approach, in which we really try to probe the physics of what is happening – ie. how are the particles accelerated, how are the inflow and outflow physically coupled, what knowledge can we extract about solid surfaces/event horizons associated with the compact object, etc.

It is fair, I think, to say that while the second approach must be the ultimate aim, it is only via the broader approach in which the significance (without the details) can be established, that the research interest in the detailed physics can be established.

4.1 Energetic and Dynamical Significance of Jets

Spectral Extent and Power of Synchrotron Component. It was already noted that transient ejection events could require a very large rate of power injection, and may be energetically significant during the outburst phase of transients (e.g. Mirabel & Rodríguez 1994). While Kaiser, Sunyaev & Spruit (2000) argued that the required energy input could be spread over a longer timescale, at least in one case, the repeated oscillation/ejection events in GRS 1915+105, this cannot be occurring (Fender & Pooley 2000).

In the persistent sources, notably the black holes in Low/Hard states, there are strong arguments that the self-absorbed synchrotron spectrum extends to at least the near-IR or optical bands (Fender 2001; Brocksopp et al. 2001; Fender et al. 2001). Since the jets are likely to be radiatively inefficient (as are AGN jets, see e.g. Celotti this volume), then the total jet power may begin to approach or even exceed the broadband X-ray luminosity, traditionally taken to be the best measure of accretion rate. In the case of the Low/Hard state transient XTE J1118+480 (which has been dynamically established to contain a black hole – McClintock et al. 2001), the ratio of total jet power to X-ray luminosity, if the synchrotron spectrum extends *only* to the near-infrared, is of order $P_{\rm J}/L_X \sim 0.01\eta^{-1}$ where η is the radiative efficiency of the jet (Fender et al. 2001). For both X-ray binaries and AGN, it is likely that $\eta < 0.1$, giving $P_{\rm J} \geq 0.1L_X$ – so even in a very conservative estimate, the jet power is at least 10% of the total accretion luminosity and cannot seriously be ignored.

Even more intriguingly, if the self-absorbed synchrotron component extends to the optical band or beyond, then comparison of the broadband radio–optical–X-ray spectra of BHCs in hard states show that the optically thin component should have a significant role to play in the X-ray band. In fact, Markoff, Falcke & Fender (2001) show that for XTE J1118 the jet can even fit almost the entire broadband spectrum, dominating ($> 90\%$) the power output of the system in the low/hard X-ray state (Fig 8). For the neutron stars the data are sparser, but Fomalont et al. (2001a,b) argue that the jet power in Sco X-1 is at least an order of magnitude greater than the observed X-ray luminosity. Furthermore, sources such as SS 433 and LS 5039 (Paredes et al. 2000) are very weak in the X-ray band by the standards of XRBs in general, yet are powerful producers of

jets and maybe even γ-rays. Perhaps we should adjust our thinking to consider that X-rays are not the only tracers of accretion power at large in the universe.

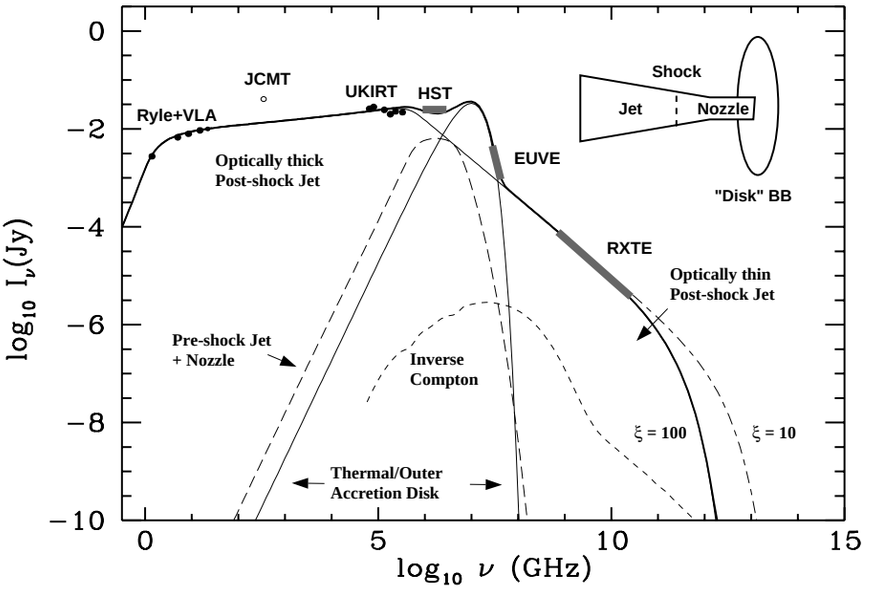

Fig. 8. Broadband radio – X-ray spectrum of the BHC XTE J1118+480 in the low/hard X-ray state, fitted by a combination of a truncated accretion disc and a jet. From Markoff et al. (2001).

Composition. The question of whether or not XRB jets are in general comprised of a normal baryonic (electrons + protons) plasma or of pairs (electrons + positrons) is important both for our concept of the flow of mass in accretion, and in estimating the energetics of the outflow. For example, in the case of the repeated oscillations in GRS 1915+105 the required power in the event that each oscillation is associated with the relativistic bulk motion of a large number of protons is much greater than if the plasma is simply electron:positron pairs (Fender & Pooley 2000).

Unfortunately, with the exception of SS 433 for which atomic emission lines have been directly observed (e.g. Margon 1984), we only observe synchrotron emission from electrons (and/or positrons) in the jets and it is not straightforward to detect the presence of protons. One ray of hope is that the use of circular polarisation measurements may shed light on the composition of jets in both AGN and XRBs (e.g. Wardle 2001), although interpretations of data are not straightforward. One XRB source, SS 433, has been detected in circular polarisation (Fender et al. 2000a), but again the relation of this source to other more 'normal' XRBs is unclear, and more measurements are needed.

Bulk Velocity. Essential to our understanding of the energetics of the outflow ($E_{\rm kinetic} \sim (\Gamma - 1) \times E_{\rm internal}$), and also for theoretical models of jet formation and their propagation through the ISM, is the 'terminal' Lorentz factor of the outflows in X-ray binaries. Only SS 433 has a well-defined and only moderately relativistic flow velocity. For other systems, even the well studied ones such as GRS 1915+105, the 'true' bulk Lorentz factor is such a sensitive function of the assumed distance (Fender et al. 1999a) and our interpretation of the data (Bodo & Ghisellini 1995) that we cannot really be certain at all of its value.

One interesting consideration is that the similarity of the correlation between X-ray and radio fluxes in the Low/Hard state sources Cyg X-1, GX 339-4 and others (Brocksopp et al. 1999; Corbel et al. 2000; Fender 2001) may naively imply that one component cannot be strongly beamed compared to the other, or the relation would be very different from source to source, depending on inclination even if they all had exactly the same velocity. Therefore it seems likely that the bulk Lorentz factor is not exceptionally high, probably < 10, although a quantitative investigation is required to tell if the available small-number statistics really are that constraining.

Jet Rotation – Angular Momentum Transport. While there is much progress in understanding the energetic significance of jets from X-ray binaries, their influence (if any) on the extraction of angular momentum from the accretion flow (a necessary but poorly-understood process) remains unclear. Observations of rotating jets from X-ray binaries, especially if the amount of angular momentum could be quantitatively estimated, would be of great significance.

Interactions with the ISM. Observations of interactions between XRB jets and the ISM are much less common than observations of interactions between AGN jets and the IGM; as a result we are left with one less diagnostic of the energetics of the outflow – in essence, the endpoint of the bulk of XRB jet power (in the form of kinetic energy) is unknown. In a few cases, interactions with the ISM have been observed – the BHCs 1E 1740.7-2942 and GRS 1758-258 in the galactic centre show AGN-like radio lobes (e.g. Mirabel 1994); SS 433 is clearly interacting with the surrounding radio nebula W50 (e.g. Dubner et al. 1998); Cir X-1 is surrounded by a radio nebula which seems to be powered by its radio jets (Stewart et al. 1993; Fender et al. 1998), and there are a few more examples.

Besides being additional clues as to the total power of jets, other intriguing possibilities exist which could be investigated by means of the jet-ISM interaction. For example, are XRB jets a source of cosmic rays? do they induce star formation?

5 Physical Processes

5.1 Jet Formation

Detailed numerical modelling of relativistic jets currently favour magnetohydrodynamic (MHD) models (e.g. Meier, Koide & Uchida 2001 and references

therein). Can we test these models with observations of X-ray binaries? perhaps in some ways we can – for example Meier (2001) takes the empirically derived association between hard X-ray states and radio emission in XRBs as some of the strongest observational evidence for MHD jet formation in geometrically thick accretion flows threaded by poloidal field lines.

5.2 Comparison of Neutron Stars and Black Holes

Are there any observed differences between the accretion:outflow coupling in BHCs and NS systems? There may be some hints – firstly, in Fender & Hendry (2000) it was established that the persistent BHCs in the Low/Hard state had approximately the same radio luminosity as the Z sources on the HB. Since the Z sources are significantly more luminous X-ray sources than the Low/Hard state BHCs, this already implied that there was some difference. In Fender & Kuulkers (2001) the ratio of peak radio to X-ray flux was compared for all reported (quasi-)simultaneous observations of X-ray transients. This ratio, or 'radio loudness' was found to be significantly higher for the BHCs. There are two obvious possible causes of this effect – either BHCs are more efficient at producing jets (extraction of energy from their deeper gravitational potentials?), or maybe the BHCs are underluminous at X-ray wavelengths, even in outburst, due to radiatively inefficient flows and their lack of a solid surface. The former explanation implies that we are probing to within the last few gravitational radii around the compact object; the latter may be considered evidence for black hole event horizons.

5.3 Particle Acceleration and a Comparison to AGN and GRBs

Part of the theme of this volume is a comparison between the physics of relativistic outflows from X-ray binaries, AGN and GRBs. I shall briefly address both of these areas here.

Particle Acceleration. Observation of the spectral index of optically thin synchrotron sources allows a direct probe of the underlying electron population. The synchrotron emission is produced by a (probably) nonthermal (power-law) distribution of relativistic (Lorentz factors possibly to ≥ 1000) electrons spiralling in a magnetic field. Typically this results in an observed power-law emission component at frequencies for which self-absorption is not important (ie. 'optically thin'), for which we can define the spectral index $\alpha = \Delta \log S_\nu / \Delta \log \nu$, i.e. the observed flux density $S_\nu \propto \nu^\alpha$ (warning : many works, especially older papers on AGN, use the reverse definition, ie. $S_\nu \propto \nu^{-\alpha}$). If the power-law distribution of electrons is described as $N(E)dE = N_0 E^{-p} dE$ (where $N(E)dE$ is the number of electrons with energy in the range E to $E + dE$, and N_0 is a constant), then the observed spectral index $\alpha = (1-p)/2$ (for a plasma in which adiabatic expansion losses dominate). Thus measurement of the optically thin

spectral index directly provides information on the distribution of relativistic electrons; typically $-1 \leq \alpha_{\rm opt.thin} \leq -0.5$, implying $2 \leq p \leq 3$.

These values are broadly consistent with those predicted for acceleration of the particles at a shock (e.g. Blandford & Eichler 1987) and are comparable to those observed in AGN. For now it seems reasonable to accept shock-accelerated power-law distributions of electrons as the origin of the observed synchrotron emission.

XRBs as Mini-AGN: 'Microquasars'. The term 'microquasar' is evocative and has been powerful in attracting public and scientific interest to the field of jets from X-ray binaries. It cannot be denied that there is some accuracy in it as a scientific expression, since in both AGN and X-ray binaries it seems that an accretion flow around a black hole (or neutron star in the case of the XRBs) results in the production of a powerful collimated outflow. There are obvious differences too, such as the supply of matter for accretion (or 'Fuel Tank') which is clearly different in the two cases, but perhaps, since the exotic physics takes place relatively close to the black hole, where the material has presumably 'forgotten' where it came from, this is not important. For example, Falcke & Biermann (1996) discuss the applicability of scaling down AGN jet models to X-ray binaries. It has often been noted that since accretion timescales might be expected to scale with mass of the accretor then processes which could never be observed from an AGN in a single human lifetime may be observed many times over by the same individual from a microquasar (e.g. Sams, Eckart & Sunyaev 1996; Mirabel & Rodríguez 1999).

Has the study of XRBs shed any light yet on the physics of AGN? this is less clear, but the prospects are good. Certainly application of many principles developed for AGN has been extremely useful in helping us to understand XRB jets without having to reinvent the wheel, and it would be nice to reciprocate. An example may be the clear relation between modes of accretion, or 'states', and the presence of radio jets in the BHC XRBs – is this related to the radio-loud:radio-quiet dichotomy in AGN?

XRBs and GRBs. GRBs and their afterglows are now widely believed to be associated with highly relativistic outflows (e.g. Sari, Piran & Halpern 1999; see also papers by Sari and Galama in these proceedings). Since they appear to be small-scale objects (compared to AGN) it is therefore natural to look for any connections with the jets of XRBs. Pugliese, Falcke & Biermann (1999) specifically discuss the possibility of a GRB from SS 433, and Portegies Zwart, Lee & Lee (1999) discuss the possibility of GRBs arising from a precessing jet in a 'gamma-ray binary'. For now the jury is out on the relevance of such comparisons and models. In at least one respect, the bulk Lorentz factor (invoked to be 100 or even more in GRBs) there seems to be a significant distinction between the physics of the outflow in the two types of object.

6 Conclusions

To conclude, the study of jets from X-ray binaries is providing much exciting data on the physics of the coupling of accretion and outlflow around both neutron stars and black holes. The latter in particular provide hope that we may be able to learn about the physics of black hole accretion in AGN, the powerhouses of the Universe, by studying XRBs. It is this author's feeling that jets will turn out to be a fairly ubiquitous characteristic of the accretion process in XRBs. In order to establish this, more observations of radio emission from the NS Atoll sources, the single largest class of XRB, are required.

Acknowledgements

I am very happy to acknowledge stimulating conversations with many of my colleagues at the SURF2000 workshop in Mykonos. I would also like to thank Sera Markoff for a careful reading of a draft of this manuscript, and Qingzhong Liu for help with XRB population numbers.

References

1. Bailyn C.D., Orosz J.A., McClintock J.E., Remillard R.A., 1995, Nature, 378, 157
2. Belloni T., Migliari S., Fender R.P., 2000, A&A, 358, L29
3. Belloni T., Mendez M., van der Klis M., Hasinger G., Lewin W.H.G., van Paradijs J., 1996, ApJ, 472, L107
4. Belloni T., Mendez M., van der Klis M., Lewin W.H.G., Dieters S., 1998, ApJ, 519, L159
5. Belloni T., Klein-Wolt M., Mendez M., van der Klis M., van Paradijs J., 2000, A&A, 355, 271
6. Bildsten L., et al., 1997, ApJ Supp. Ser., 113, 367
7. Blandford R., Eichler, 1987, Physics Reports 154, 1
8. Bodo G., Ghisellini G., 1995, ApJ, 441, L69
9. Bradshaw C.F., Fomalont E.B., Geldzahler B.J., 1999, ApJ, 512, L121
10. Brocksopp C., 2000, PhD thesis, University of Sussex, UK
11. Brocksopp C., Fender R.P., Larionov V., Lyuty V.M., Tarasov A.E., Pooley G.G., Paciesas W.S., Roche P., 1999, MNRAS, 309, 1063
12. Brocksopp C., Jonker P.G., Fender R.P., Groot P.J., van der Klis M., Tingay S.J., 2001, MNRAS, 323, 517
13. Campana S., Colpi M., Mereghetti S., Stella L., Tavani M., 1998, A&ARev, 8, 279
14. Charles P.A., 1998, In : Theory of Black Hole Accretion Disks, Eds M. A. Abramowicz, G. Bjornsson, and J. E. Pringle. Cambridge University Press, 1998., p.1
15. Chen W., Shrader C.R., Livio M., 1997, ApJ, 491, 312
16. Corbel S., Fender R.P., Tzioumis A.K., Nowak M., McIntyre V., Durouchoux P., Sood R., 2000, A&A, 359, 251
17. Corbel S. et al., 2001, ApJ, 554, 43
18. Dhawan V., Mirabel I.F., Rodríguez L.F., 2000, ApJ, 543, 373
19. Dubner G.M., Holdaway M., Goss W.M., Mirabel I.F., 1998, AJ, 116, 1842

20. Eikenberry S.S., Matthews K., Morgan E.H., Remillard R.A., Nelson R.W., 1998, ApJ, 494, L61
21. Eikenberry S.S., Matthews K., Muno M., Blanco P.R., Morgan E.H., Remillard R.A., 2000, ApJ, 532, L33
22. Esin A.A., McClintock J.E., Narayan R., 1997, ApJ, 489, 865
23. Falcke H., Biermann P.L., 1996, A&A, 308, 321
24. Fender R.P., 2001, MNRAS, 322, 31
25. Fender R.P., Hendry M.A., 2000, MNRAS, 317, 1
26. Fender R.P., Pooley G.G., 2000, MNRAS, 318, L1
27. Fender R.P., Kuulkers E., 2001, MNRAS, 324, 923
28. Fender R.P., Pooley G.G., Brocksopp C., Newell S.J., 1997, MNRAS, 290, L65
29. Fender, R., Spencer, R., Tzioumis, T., Wu, K., van der Klis, M., van Paradijs J., Johnston H., 1998, ApJ, 506, L21
30. Fender, R.P., Garrington, S.T., McKay, D.J., Muxlow, T.W.B., Pooley, G.G., Spencer, R.E., Stirling, A.M., Waltman, E.B., 1999a, MNRAS, 304, 865
31. Fender R. et al., 1999b, ApJ, 519, L165
32. Fender R., Rayner D., Norris R., Sault R.J., Pooley G., 2000a, ApJ, 530, L29
33. Fender R.P., Pooley G.G., Durouchoux P., Tilanus R.P.J., Brocksopp C., 2000b, MNRAS, 312, 853
34. Fender R.P., Hjellming R.M., Tilanus R.P.J., Pooley G.G., Deane J.R., Ogley R.N., Spencer R.E., 2001, MNRAS, 322, L23
35. Feretti L., et al., 2001, Proceedings of the 5th EVN Symposium, Eds. J. Conway, A. Polatidis, R. Booth, Onsala Observatory, Sweden (June 2000), in press (**astro-ph/0009348**)
36. Fomalont E.B., Geldzahler B.J., Bradshaw C.F., 2001a, ApJ, 553, L27
37. Fomalont E.B., Geldzahler B.J., Bradshaw C.F., 2001b, ApJ, in press (**astro-ph/0104372**)
38. Garcia M.R., Grindlay J.E., Molnar L.A., Stella L., White N.E., Seaquist E.R., 1988, ApJ, 328, 552
39. Geldzahler B.J. et al., 1983, ApJ, 273, L65
40. Gomez J.-L., Marscher A.P., Alberdi A., Jorstad S.G., Garcia-Miro C., 2000, Science, 289, 2317
41. Hasinger G., van der Klis M., 1989, A&A, 225, 79
42. Hjellming R.M., Johnston K.J., 1981a, Nature, 290, 100
43. Hjellming R.M., Johnston K.J., 1981b, ApJ, 246, L141
44. Hjellming, R.M., Han, X., 1995, Radio properties of X-ray binaries. In : Lewin, W.H.G., van Paradijs, J., van der Heuvel, E.P.J. (Eds.), X-ray binaries, Camridge University Press, Cambridge, 308–330
45. Hjellming, R.M., Rupen, M.P., 1995, Nature, 375, 464
46. Homan J., Wijnands R., van der Klis M., Belloni T., van Paradijs J., Klein-Wolt M., Fender R., Mendez M., 2001, ApJS, 132, 377
47. Janiuk A., Czerny B., Zycki P.T., 2000, MNRAS, 318, 180
48. Kaiser C.R., Sunyaev R., Spruit H.C., 2000, A&A, 356, 975
49. Klein-Wolt M., Fender R.P., Pooley G.G., Bellpni T.M., Morgan E.H., Migliari S., van der Klis M., 2001, MNRAS, submitted
50. Kuulkers E., Fender R.P., Spencer R.E., Davis R.J., Morison I., 1999, MNRAS, 36, 919
51. Liu Q.Z., van Paradijs J., van den Heuvel E.P.J., 2000, A&AS, 147, 25
52. Liu Q.Z., van Paradijs J., van den Heuvel E.P.J., 2001, A&A, 368, 1021
53. Livio M., 1999, Physics Reports, 311, 225

54. McClintock J.E., Garcia M.R., Caldwell N., Falco E.E., Garnavich P.M., Zhao P., 2001, ApJ, in press
55. Markoff S., Falcke H., Fender R.P., 2001, A&A, 372, L25
56. Margon B., 1984, ARA&A, 22, 507
57. Martí J., Mirabel I.F., Rodríguez L.F., Chaty S., 1998, A&A, 332, L45
58. Meier D.L., 2001, ApJ, 548, L9
59. Meier D.L., Koide S., Uchida Y., 2001, Science, 291, 84
60. Mendez M., Belloni T., van der Klis M., 1998, ApJ, 499, L187
61. Meier D.L., 2001, ApJ, 548, L9
62. Mioduszewski A.J., Rupen M.P., Hjellming R.M., Pooley G.G., Waltman E.B., 2001, ApJ, in press
63. Mirabel I.F., 1994, ApJS, 92, 369
64. Mirabel, I.F., Rodríguez, L.F, 1994, Nature, 371, 46
65. Mirabel, I.F., Rodríguez, L.F, 1999, ARA&A, 37, 409
66. Mirabel I.F., Rodríguez L.F., Cordier B., Paul J., Lebrun F., 1992, Nature, 358, 215
67. Mirabel I.F., Dhawan V., Chaty S., Rodríguez L.F., Martí J., Robinson C.R., Swank J., Geballe T., 1998, A&A, 330, L9
68. Naik S., Rao A.R., 2000, A&A, 362, 691
69. Paragi Z., Fejes I., Vermeulen R.C., Schilizzi R.T., Spencer R.E., Stirling A.M., 2001, in: Schilizzi R.T., Vogel S., Paresce F., Elvis M. (eds.) Proc. IAU Symposium 205 "Galaxies and their constituents at the highest angular resolutions", ASP, p. 112
70. Paredes J.M., Martí J., Ribo M., Massi M., 2000, Science, 288, 2340
71. Penninx W., 1989, in Proc: The 23rd ESLAB Symposium on Two Topics in X Ray Astronomy. Volume 1: X Ray Binaries, p. 185
72. Penninx W., Lewin W.H.G., Zijlstra A.A., Mitsuda K., van Paradijs J., 1988, Nature, 336, 146
73. Pooley G.G., Fender R.P., 1997, MNRAS, 292, 925
74. Pooley G.G., Fender R.P., Brocksopp C., 1999, MNRAS, 302, L1
75. Portegies Zwart S.F., Lee C.H., Lee H.K., 1999, A&AS, 138, 503
76. Poutanen J., 1998, In : Abramowicz, M. A., Björnsson, G., Pringle, J. E. (Eds), Theory of Black Hole Accretion Discs, Cambridge Contemporary Astrophysics, CUP, 1998, p.100
77. Pugliese G., Falcke H., Biermann P.L., 1999, A&A, 344, L37
78. Rees M.J., 1966, Nature, 211, 468
79. Reig P., Mendez M., van der Klis M., Ford E.C., 2000, ApJ, 530, 916
80. Rodríguez L.F., Mirabel I.F., Martí J., 1992, ApJ, 401, L15
81. Sams B., Eckart A., Sunyaev R., 1996, Nature, 382, 47
82. Sari R., Piran T., Halpern J.P., 1999, ApJ, 519, L17
83. Shirey R.E., Bradt H.V., Levine A.M., 1999, ApJ, 517, 472
84. Spencer R.E., 1979, Nature, 282, 483
85. Stewart, R.T., Caswell J.L., Haynes, R.F., Nelson, G.J., 1993, MNRAS, 261, 593
86. Stirling A., Spencer R., Garrett M., 1998, New Astronomy Reviews, 42, 657
87. Stirling A.M., Spencer R.E., de la Force C.J., Garrett M.A., Fender R.P., Ogley R.N., 2001, MNRAS, submitted
88. Tananbaum H., Gursky H., Kellogg E., Giacconi R., Jones C., 1972, ApJ, 177, L5
89. Tingay, S.J. et al., 1995, Nature, 374, 141
90. van der Klis, M., 1995, In : Lewin, W. H. G., van Paradijs, J., van der Heuvel, E. P. J. (Eds.), X-ray binaries, Camridge University Press, Cambridge, 252

91. van Paradijs, J., 1995, In : Lewin, W. H. G., van Paradijs, J., van der Heuvel, E. P. J. (Eds.), X-ray binaries, Camridge University Press, Cambridge, 536
92. Wardle J.F.C., 2001, In Proc 'Particles and Fields in Radio Galaxies', Eds. Robert A. Laing and Katherine M. Blundell, ASP Conference Series, in press (**astro-ph/0011515**)
93. Zensus J.A., Pearson T.J. (Eds), 1987, Superluminal Radio Sources, Cambridge University Press
94. Zhang S.N., Mirabel I.F., Harmon B.A., Kroeger R.A., Rodríguez L.F., Hjellming R.M., Rupen M.P., 1997, In Proc. Fourth Compton Symposium, C.D. Dermer, M.S. Strickman & J.D. Kurfess (Eds), AIP Conf. Proc. 410, p. 141

Gamma-Ray Bursts: The Afterglow Revolution

Titus J. Galama[1] and Re'em Sari[2]

[1] Astronomy 105-24, California Institute of Technology, Pasadena CA 91125, USA
[2] Theoretical Astrophysics 130-33, California Institute of Technology, Pasadena CA 91125, USA

Abstract. GRBs were discovered with the Vela satellites, whose main purpose was to verify compliance with the 1963 Limited Nuclear Test Ban Treaty. Since their discovery these events, which emit the bulk of their energy in the $0.1 - 1.0$ MeV range, and whose durations span milliseconds to tens of minutes, posed one of the great unsolved problems in astrophysics. GRBs are formed in extreme relativistic outflows and provide important information about highly relativistic acceleration mechanisms. Until 1997, no counterparts (quiescent as well as transient) could be found and observations did not provide a direct measurement of their distance. The breakthrough came in early 1997, when the Wide Field Cameras aboard the Italian-Dutch BeppoSAX satellite allowed rapid and accurate localization of GRBs. Follow-up on these positions resulted in the discovery of X-ray, optical and radio afterglows. These observations revealed that GRBs come from 'cosmological' distances, and that they are by far the most luminous photon sources in the Universe, with peak luminosities in γ rays up to 10^{52} erg/s, and total energy budgets up to several times 10^{53-54} erg (for assumed isotropic emission). Evidence is accumulating, however, that GRB outflow is collimated in the form of jets and when corrected for the geometry of the outflow the energies of GRBs appear to cluster around 5×10^{50} ergs- very comparable to that of supernovae. GRBs are rare phenomena with an overall rate about 2000 times smaller than that of supernovae. Indirect evidence in the last several years shows that a fraction of GRBs may be related to a peculiar type of supernova explosions. Theoretical work has shown that these supernovae most likely mark the birth events of stellar mass black holes as the final products of the evolution of very massive stars. A fundamental question is whether there are also other processes that can drive such an engine, for example the coalescence of a double neutron-star system. Finally, the expectation is that one can use the enormous optical and infrared luminosities of the afterglows of GRBs to probe the Universe out to very large redshifts, beyond what is possible using supernovae or quasars. This would open an entirely new and exciting field of astrophysics and cosmology.

1 Introduction

In this article we will attempt to provide an overview of the current status of Gamma-Ray Burst (GRB) research. In particular, we focus on the observational and theoretical understanding that has resulted from the identification of counterparts to GRBs, i.e. the discovery of long-lived afterglows.

The structure of the review will be as follows. In Sect. 2 we will provide a brief discussion of GRB properties, and some history of GRB research. In Sect. 3 and 4 we present the, by now, 'standard' model for interpretation of the gamma-ray burst and of the afterglow emission: the fireball plus relativistic blast-wave

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 123–168, 2002.

model. In Sect. 5 we discuss the afterglow revolution, which revealed the extreme energetics of GRBs and confirmed the highly relativistic nature of these sources. In Sect. 6 and 7 we discuss the theory of collimated (jetted) outflow and present the observational evidence for collimation in GRBs. In Sect. 9 we present the physics of the reverse shock and discuss the extreme optical brightnesses that GRB afterglows may show at very early times. In Sect. 10 we discuss possible progenitor systems of GRBs and the observational constraints we have on those. The huge luminosities at early times may very well be used to study the very high-redshift Universe ($z > 5$), a topic that will be touched upon in Sect. 11.

2 γ-Ray Burst Properties

2.1 Vela Satellites

The Limited Nuclear Test Ban Treaty prohibits nuclear weapons tests 'or any other nuclear explosion' in the atmosphere, in outer space, and under water. The Vela satellites were designed such that they could verify compliance with the treaty by detecting the γ–rays from nuclear tests outside the Earth's atmosphere. Vela 5A, 5B, 6A and 6B each carried six 10 cm^3 CsI scintillation counters; they could detect photons in the 0.2-1.0 MeV (Vela 5) and 0.3-1.5 MeV (Vela 6) energy range (see Fig. 1).

A search for γ-ray bursts was started by R. Klebesabel because of the prediction that γ-ray emission would be observable during the initial stages of super-

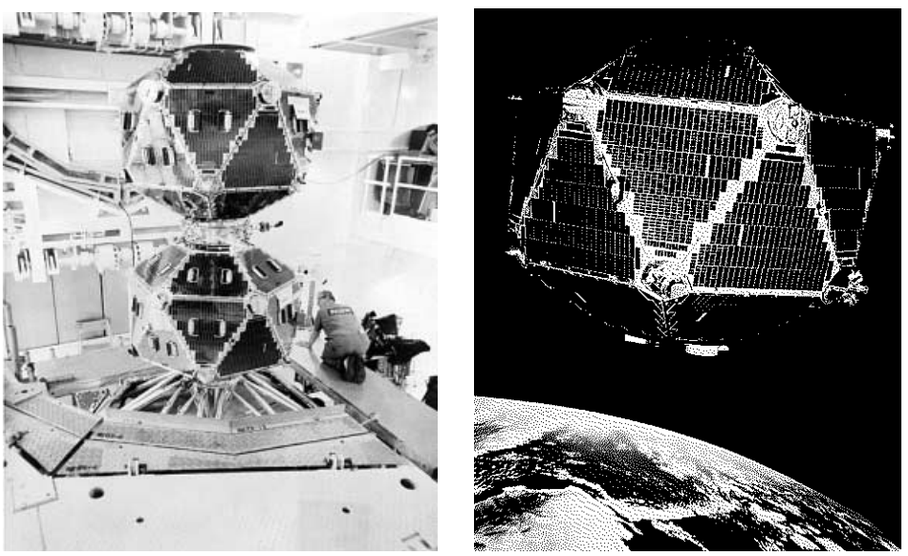

Fig. 1. *Left figure: Vela-5A and 5B Satellites in the Clean Room. Right figure: animation of Vela-5B in low Earth orbit (courtesy of Los Alamos National Laboratory).*

nova explosions [1]. No indications for such a phenomenon were found. It was in 1969 that, embedded in Vela spacecraft data from 1967, a γ-ray burst was found. At that time the Sun could not be excluded as a source. With the launch of a new generation of Vela satellites (the Vela 6) sufficient timing accuracy made it possible to exclude the Sun as the source of these events. Once about 16 events were found, Klebesabel, Strong and Olson published a paper announcing the discovery of cosmic γ-ray bursts [2]. The original July 1967 event is not in that paper because it could have come from the Sun (although its characteristics are like a GRB and it is now considered the oldest known GRB)[1]. A time history of the 1967 event can be found in Strong and Klebesabel [3].

2.2 Light Curves and Spectra of GRBs

The Burst And Transient Source Experiment (BATSE) on board the Compton Gamma-Ray Observatory (CGRO; see Fig. 2) observed about one GRB a day. The CGRO provided a wealth of information on GRB light curves and spectra; here we summarize some of the most important results. A more extensive discussion on temporal properties and spectra of GRBs can be found in the review by Fishman and Meegan [4].

- GRBs have durations ranging from milliseconds to $\sim 10^3$ s. Their time histories display a great diversity of structure, with single or multiple peaks, with smooth profiles, profiles with sub-peaks, profiles with well separated peaks and profiles with overlapping peaks and spikes (see Fig. 3).
- GRB light curves show rapid variability, on time scales sometimes less than a millisecond [5,6]. As light travels a distance $L = ct = 3 \cdot 10^7$ cm in one millisecond, about 10 neutron star radii, it is generally believed that GRBs originate from compact objects, such as neutron stars (NS) or black holes (BH).
- Several attempts have been made to categorize GRB time histories. No other observational parameters appear to be correlated with temporal morphologies (e.g., [4]), i.e., there is little morphological evidence for distinct burst classes within the GRBs. There is one exception: the distribution of burst durations is bimodal and separates GRBs into two classes, the short events (<2s) and the longer ones (>2s) ([7]; see Fig. 4). The duration appears to be anticorrelated with spectral hardness: short bursts are predominantly harder than long ones [7].
- Norris et al. [9] find evidence for time dilation by comparison of samples of bright and dim BATSE GRBs. The centroids and widths of the duration distribution for the dim sample are scaled by a factor of two relative to the bright sample. They interpret this as a result of cosmological redshift: the dim bursts are, on average, located at larger distance (i.e., redshifts) than the

[1] It is often believed that the publication was delayed until 1973 because the data were classified. In fact it was the better timing of GRBs which was needed to rule out the Sun.

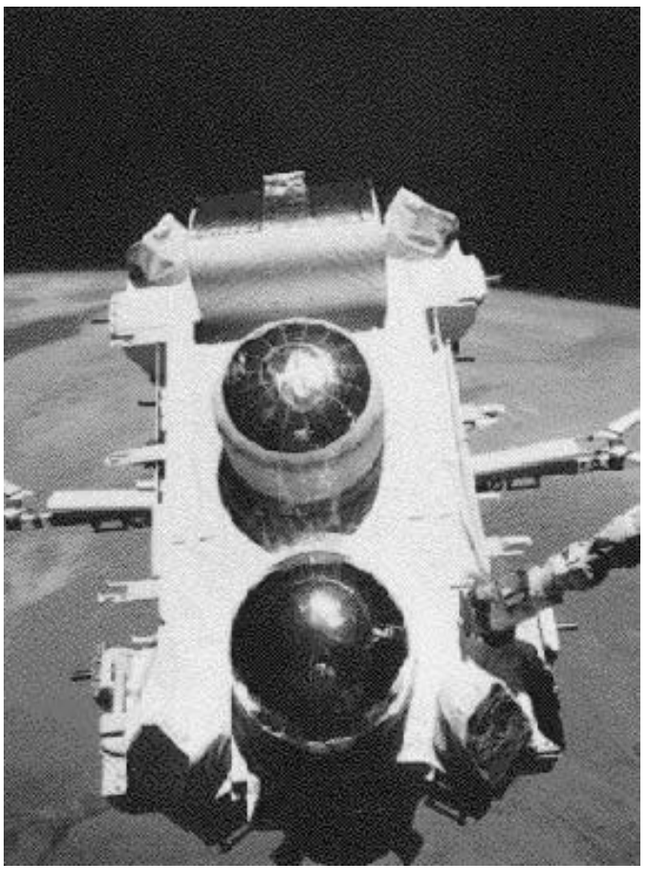

Fig. 2. *The Compton Gamma Ray Observatory (CGRO) above western Africa just prior to its release from the Space Shuttle into orbit. Visible are the solar panels and the four CGRO experiments: the large round domes of the EGRET (bottom) and COMPTEL (center) experiments and four of the eight BATSE detectors located at the corners of the satellite are visible. The OSSE experiment housing is visible just above the COMPTEL dome (courtesy of NASA).*

bright bursts. The sources of dimmer bursts would lie at redshifts of around 2.

- High-energy emission is a unique feature of GRBs. Spectral measurements extend from a few keV to $\sim$ GeV (see e.g., Fig. 5). The continuum spectra of GRBs are very broad and hard; most of the power is emitted above 50 keV. GRB spectra are well described by an empirical function, the so called Band function [10]. This function consists of a low- and high-energy power law, smoothly joined by an exponential turnover. It has three parameters, the peak energy, and the low- and high-energy photon index (for details see [10]).

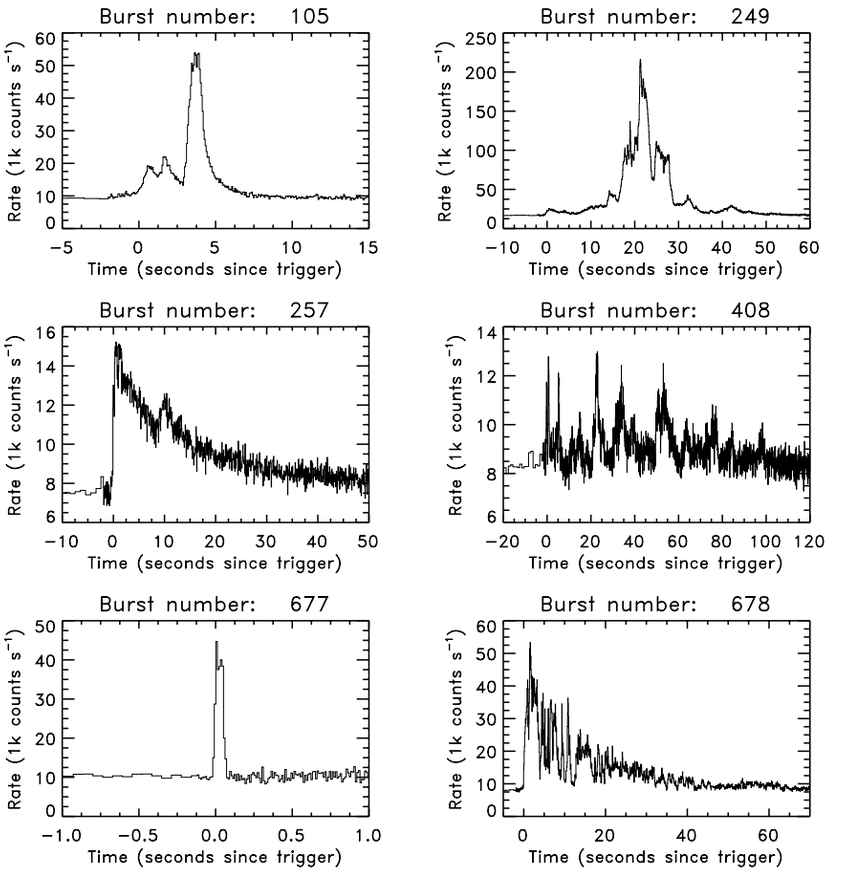

Fig. 3. *A sample of GRB light curves (25-2000 keV) observed by BATSE. The light curves display a large variety in time profiles, duration, and intensities. Burst trigger 105 is an example of a burst with smooth, well defined peaks. Trigger 257 is a typical example of a Fast Rise, Exponential Decay (FRED) light curve. Trigger 408 is an example of a complex, chaotic and spiky burst. Trigger 677 is a GRB with very short duration, and belongs to the sub-class of short GRBs (< 2 sec) (courtesy of the BATSE team).*

- Delayed very high-energy γ rays have been observed from GRBs with EGRET on board the CGRO [11,12]. In the case of GRB 940217, 200 MeV-20 GeV photons were observed with EGRET up to 90 minutes after the event onset [11], much longer than the duration of the lower photon energies of the GRB itself ($\sim$ 200 seconds; 30-2000 keV).
- Ford et al. [13] investigated the evolution of the peak energy for long and bright GRBs and found that the peak energy decreases with time. Liang and Kargatis [14] found that the peak energy decreases exponentially with the pho-

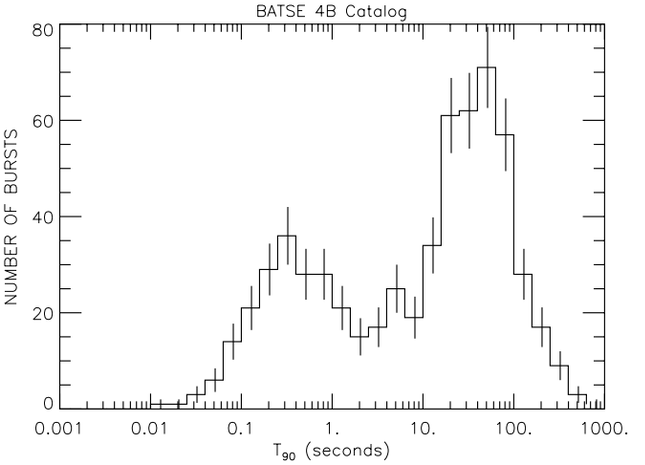

Fig. 4. *Distribution of T_{90} for BATSE bursts from the 4B catalog [8]. T_{90} is defined as the time during which the cumulative counts increase from 5% to 95% [7].*

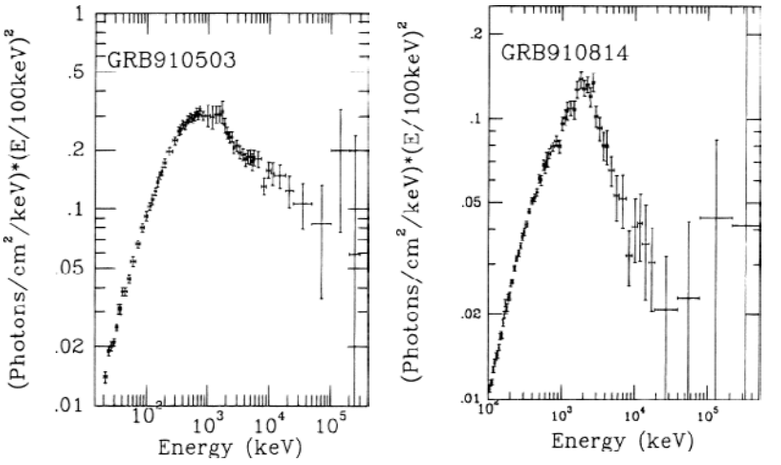

Fig. 5. *Spectra of two bright GRBs (GRB 910503, 20 keV to 300 MeV; GRB 910814, 103 keV to 500 MeV) from BATSE, COMPTEL and OSSE data (from [19]).*

ton fluence. This 'hard-to-soft' spectral evolution (see also [15]) is also found within individual pulses of GRBs [13].

- Time histories of GRBs are different for different energy bands. At higher energies the overall burst duration as well as the rise and fall time scales of pulses are shorter than those at lower energies [16]. Typically, the low-energy emission in GRBs persists longer than the high-energy emission.

2.3 Distribution on the Sky and in Space

Prior to the launch of the CGRO, in 1991, various experiments discovered, that the burst population is approximately uniform in space. Detector sensitivities were still too small to detect an 'edge' to the distribution of burst sources [17,18].

The space distribution of GRB sources is related to their apparent flux distribution, as can be seen from the following simplified argument. We first assume that bursts are "standard candles", i.e., they emit the same amount of luminosity at the same wavelengths. A burst at a distance d_0 is observed with a peak flux P_0, $P_0 \propto d_0^{-2}$. The cumulative rate of bursts, with peak fluxes P exceeding P_0, $R(\geq P_0)$, is proportional to the volume V_0 of space in which bursts with observed peak fluxes $P \geq P_0$ can be observed. Hence it is proportional to d_0^3 and, by $P_0 \propto d_0^{-2}$, proportional to $P_0^{-3/2}$, yielding a $\log R(\geq P_0)$-$\log P_0$ curve with a slope of $-3/2$. Now, even if the luminosities have a broad distribution, they are some (or integral) of several "standard candle" population, and the some of curves with $-3/2$ slope, is still a curve with $-3/2$ slope. In the 1980's it was observed that bright bursts follow this slope of –3/2. However, balloon experiments, with very large area detectors, found that the weaker bursts deviate from the $-3/2$ slope [20]. Relatively fewer faint sources than expected from a homogeneous distribution of bursts were seen. It was not clear at that time whether this deviation was the result of selection biases or a really existing effect.

The BATSE Surprise (1991). Astronomical objects in the vicinity of the Sun (closer than the Galactic disk thickness) show a uniform distribution in space and are isotropic on the sky. The general notion in the GRB community, based on these observational facts, was that GRBs were produced by neutron stars and that we observed the nearby ones in the solar neighborhood. This notion was reinforced by the alleged detection of cyclotron lines (reflecting a magnetic field $\sim 10^{12}$ Gauss) in the spectra of several GRBs with the Venera and Ginga satellites [21,22,23], and by emission features around 400 keV, which were interpreted as gravitationally redshifted 511 keV annihilation lines [21]. However, in spite of very extensive searches, BATSE has not found any such spectral features [24,25]. It was believed that a sensitive instrument like BATSE would see the Milky Way of bursters. The big BATSE surprise [26] was that even faint GRBs are distributed isotropically on the sky (see Fig. 6). In addition, the fainter bursts showed a turnover in the cumulative brightness distribution (see Fig. 7), i.e., there is a distinct dearth of very weak GRBs. BATSE also sees no clustering of bursts on small or large angular scales, i.e., GRBs are not associated with concentrations of mass on any distance scale (e.g the Galactic disc, nearby clusters of stars, the Large Magellanic Cloud, nearby galaxies like M31 or clusters of galaxies like Virgo). The simplest explanation for these observations is that we are at the center of a spherically symmetric distribution of GRBs and that we are observing the 'edge' of this distribution.

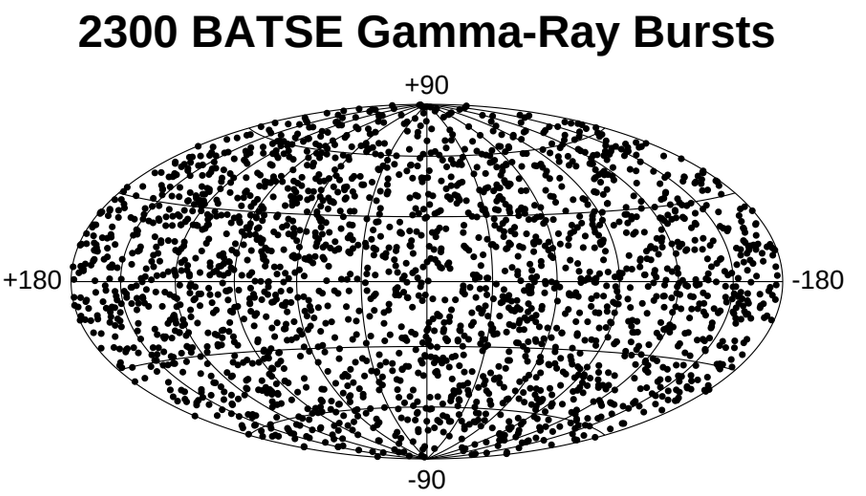

Fig. 6. *Angular distribution of BATSE bursts in Galactic coordinates.*

2.4 The 'Great' Debate

The BATSE observations (see the previous Sect. 2.3) excluded that GRBs come from a galactic disk population [26]; a Galactic disc population of burst sources would either be homogeneous, at distance scales less than several hundred pc, or anisotropic, at distances greater than several hundred pc. Naturally these results led to the conclusion that GRBs originate from 'cosmological' (Gpc) distances [8,28,29,30]. There was, however, a countervailing view that GRBs originate in a very large halo around our galaxy [31,32,33]. As a result, the discussion on the nature of GRB sources focused on their distance scale.

Galactic Halo. All known Galactic objects are strongly concentrated to the Galactic center. In order to have no measurable anisotropy of a Galactic population of bursts, the objects should populate an extended halo with a very large core radius, $R \geq 100$ kpc, so that the offset of the Sun from the center of the Galaxy does not show up. The lack of any observable concentration of bursts towards the nearby galaxy M31 gives some further constraints on the size of a Galactic burster halo. The halo radius was estimated to lie between 100 to 300 kpc [34,35,36]. This would imply that bursts should populate a yet unobserved very extended hypothetical halo. It was suggested that such an extended halo might perhaps be created by injection of high-velocity neutron stars from the Galactic plane, escaping the gravitational field of the galaxy (e.g., [37]). Assuming a distance $d = 100$ kpc for a Galactic burst we find a luminosity $L_{\rm GRB} \sim 10^{42}{\rm erg\,s^{-1}}$ (for a typical GRB flux, $f_{\rm GRB} = 10^{-6}{\rm erg\,cm^{-2}s^{-1}}$, and assuming isotropic emission).

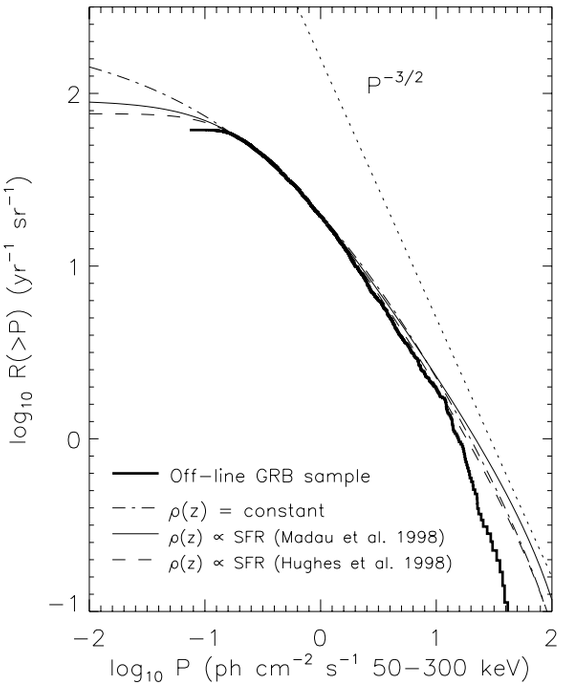

Fig. 7. *The cumulative peak flux distribution of 'triggered' BATSE bursts combined with the 'non-triggered' BATSE bursts found in 6 years of archival BATSE data (solid line; units are bursts* yr^{-1} sr^{-1}*; from [39]). Shown are the best-fit cosmological models with power-law luminosity distributions for: a co-moving burst rate that is independent of redshift (dot-dashed line), and for co-moving burst rates that follow the star-formation rate as determined by Madau et al. [40] and Hughes et al. [41]. The slope* $-3/2$ *for a uniform distribution of sources is also indicated.*

Cosmological. As far as we know, on much larger distance scales, several Gpc, all objects are distributed roughly isotropically and uniformly. A cosmological distribution of sources would reveal a natural deficiency of weaker bursts by relativistic effects that affect the weaker (on average at greater distances) bursts [8,28,30]. Because of cosmological time dilation the count rate is lower by a factor $1 + z$. Also, space is not Euclidian, which becomes noticeable at larger scales, and the observed spectrum is redshifted by a factor $1 + z$, i.e., a different part of the spectrum is observed (e.g., [38]). For a typical GRB flux, $f_{\rm GRB} = 10^{-6}{\rm erg\,cm^{-2}s^{-1}}$, and a cosmological distance of $d = 3$ Gpc we find a luminosity $L_{\rm GRB} \sim 10^{51}{\rm erg\,s^{-1}}$.

2.5 Counterparts

The lack of knowledge about the nature of GRBs led to a lot of speculation and an enormous number of proposed models. Some examples are the merger

of a double neutron star (NS-NS) or of a neutron star and black hole (NS-BH) binary, failed supernovae, asteroids or comets falling into black holes or neutron stars, processes in the core of active Galactic nuclei etc. (see [42] for a list of more than 100 models of GRBs published before 1992).

As mentioned above, the γ-ray properties of GRBs did not provide unambiguous distance clues, and it was generally agreed that settling the debate on the GRB origin required the identification of GRB counterparts at other wavelengths (e.g., [43]). Identification of a known type of object associated with a GRB would immediately reveal the distance scales and thereby greatly constrain theoretical models of GRB production. However, the generally large positional errors of bursts and the fact that accurate positions were usually obtained only long after the event, made searches for counterparts difficult. Historically, three strategies were employed for counterpart searches. Deep searches for quiescent counterparts to accurately localized events, i.e., low-energy emission long after the burst, have been made and reveal that quiescent counterparts are very faint, i.e., not detected, at all energies [43]. For example, in radio at 2, 6, and 20 cm, upper limits of $\sim$ 100–800 μJy on quiescent counterparts for ten small GRBs error regions were given [44]. Another strategy has been to search for flaring counterparts in simultaneous wide-field monitoring experiments in the hope to have a GRB included in the field of view of the instrument at the moment of the event. A third counterpart search technique is to use the GRB detection as a trigger to point the telescope in the appropriate direction and search for a flaring and/or fading counterpart on time scales as long as, and much longer than the burst itself; it is this third technique that proved to be successful (see Sect. 5). Finally, optical emission simultaneous with a GRB was discovered [45] (the strategy to search for flaring counterparts proved successful too).

3 GRB Theory – the Generic Picture

The exact mechanism leading to the phenomenon of GRBs is yet unknown, and is a matter of debate, as we will discuss in Sect. 10. Despite this fact, some basic characteristics are well understood. Below, we show how the observed spectra, energies and timescales of GRBs have lead to a generic model, the so called fireball shock model that is almost independent of knowledge about the unknown 'inner engine'.

The extreme characteristics of GRBs, i.e. the observed large energies and short timescales, lead to a paradox, the so called 'compactness problem'. An energy of 10^{52}erg is released within a variability time $\delta T \sim 0.1$s in the form of photons of about 1 MeV. This translates into a huge number of $N = 10^{56}$ photons. If we now assume that the energy is released in a small volume of linear dimensions $R \leq c\delta T \sim 10^9$cm (which is naively required by the variability timescale), then the optical depth to pair creation would be the number of photons per unit area, multiplied by the Thomson cross section σ_T or

$$\tau \sim \sigma_T \frac{N}{4\pi R^2} \sim 3 \times 10^{11} \gg 1.$$

But, if that were true, such a large optical depth implies that all the photons will have created pairs and thermalized. However, the observed spectrum of GRBs, as shown in the previous section is highly non-thermal!

The only known solution to the 'compactness problem' is relativistic motion [8,46]. These effects were considered in detail in [47,48,49]. A critical review of these as well as some new limits are given by Lithwick and Sari [50]. If the emission site is moving relativistically, with a Lorentz factor γ, toward the observer, then the optical depth is reduced, compared to the stationary estimate, due to two effects. First, the size of the source can be larger by a factor of γ^2. This will still produce variability over a short time scale given by $\delta T = R/\gamma^2 c$ since not all of the source is seen as the radiation for a relativistically moving object is beamed (see figure 8). Second, the photons in the local frame are softer by a factor of γ, and therefore only a small fraction of them, the ones at the high-energy tail of the GRB spectrum, have enough energy to create pairs. The combination of these two effects reduces the optical depth by a factor of $\sim \gamma^{6.5}$, where the exact power depends on the GRB spectrum (see [50]). Therefore, the optical depth is reduced below unity, and the 'compactness problem' is solved, if the Lorentz factor is larger than about a hundred.

This solution to the compactness problem led to a three stage generic scenario for GRBs. First, a compact source releases about 10^{52} erg, in a small volume of space and on a short time scale. This large concentration of energy expands due to its own pressure. If the rest mass that contaminates the site is not too large, $\leq 10^{-5} M_{\odot}$ (the requirement of a small baryonic load), this will result in relativistic expansion with $\gamma > 100$. Finally, at a large enough radius, the kinetic energy (bulk motion) of the expanding material is converted to internal energy and radiated, mainly in γ-rays. At this stage the system is optically thin and high energy photons can escape. We now discuss this third stage in some detail.

3.1 Internal vs. External Shocks

Assume a flow carrying 10^{52} erg as kinetic energy. In order for this to produce photons, the kinetic energy must be converted back into internal energy and radiated away. The flow must therefore, at least partially, slow down. Two scenarios were proposed for this deceleration: external shocks [51] and internal shocks [52,53]. In the external shocks scenario, the relativistic material is running into some (external) ambient medium, possibly the interstellar medium (ISM) or a stellar wind that was emitted earlier by the progenitor. In the internal-shocks scenario the inner engine is assumed to emit an irregular flow, that consists of many shells, that travel with a variety of Lorentz factors and therefore collide with one another and thermalize part of their kinetic energy.

The property that proved to be very useful in constraining these two possibilities is the variability observed in many of the bursts. In the external-shocks scenario, this variability is attributed to irregularities in the surrounding medium, e.g., clouds. Each time the ejecta runs into a higher density environment, it produces a peak in the emission. In the internal shocks scenario, the source has to emit many shells, and when two of them collide a peak in the emission is

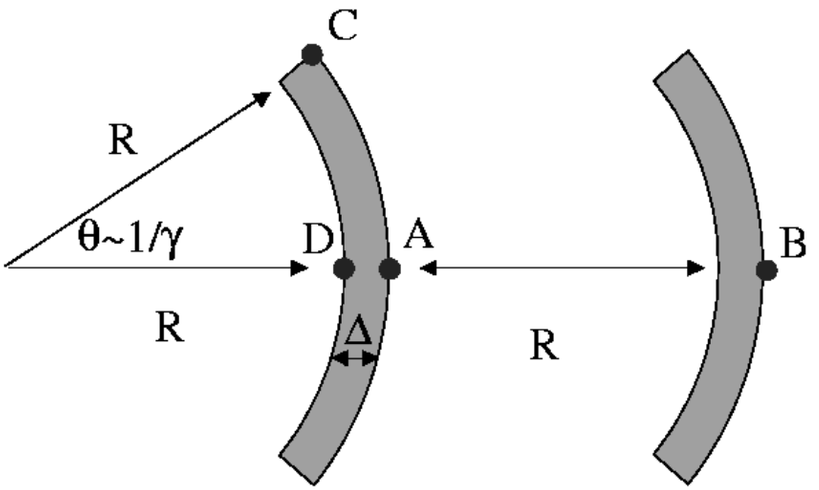

Fig. 8. *Timescales from an expanding relativistic fireball. The gray area represents the observed section of the fireball that can be seen by an observer located far to the right. The angular opening of that section is* $1/\gamma$ *due to relativistic beaming. Consider the 4 photons emitted at points A, B, C, and D. Photons A, C and D where emitted simultaneously, but photon A will arrive at the observer first, since it is closer to the observer. The arrival-time delay of photons C and D with respect to photon A is simply given by the extra distance they have to travel. Therefore* $\delta T_{C-A} = R(1-\cos\theta)/c = R/2\gamma^2 c$*, and* $\delta T_{D-A} = \Delta/c \sim R/\gamma^2 c$*, where we have used the fact that relativistic dynamics of fireballs imply* $\Delta \sim R/\gamma^2$*. Finally, photon B was emitted long after photon A (about a time* R/c *later than photon A), however, it is much closer to the observer, resulting in* $\delta T_{B-A} = R/2\gamma^2 c$*. All three timescales lead to the expression* $R/\gamma^2 c$*. A short observed variability time scale can therefore be obtained even for large radius, if the Lorentz factor is sufficiently high. The naive estimate of* $R \leq c\delta T$ *is, therefore, to be replaced by* $R \leq \gamma^2 c\delta T$*.*

produced. External shocks thus require a complicated surrounding with a relatively simple source that explodes once, while internal shocks require a more complicated source that will explode many times to produce several shells. Due to these very different requirements on the source, the question of internal or external shocks is of a fundamental importance in understanding the nature of the phenomenon.

The size of the clouds that the ejecta runs into, in the external-shocks scenario, has to be very small in order to produce peaks that are narrower than the duration of the burst [54]. Sari & Piran [55] gave the following argument. The size of the clouds has to be smaller than $R/N\gamma$ to produce peaks that are narrower by a factor of N than the duration of the burst. The number of clouds should be smaller than N otherwise pulses arriving from different clouds will overlap and the amplitude of the variability will be reduced. Finally, the observable area of the ejecta, due to relativistic beaming is $(R/\gamma)^2$. The maximal

efficiency of the external shocks scenario is therefore given by

$$\frac{\text{cloud area} \times \text{number of clouds}}{\text{observed shell area}} \leq \frac{1}{N} \sim 1\%. \quad (1)$$

Since in many bursts $N > 100$, external shocks have a severe efficiency problem in producing highly variable bursts. Also other predictions of external shocks are inconsistent with the observed temporal profile [56]. Moreover, the density ratio between the clouds and the surroundings has to be huge, of the order of $\gamma N^2 \sim 10^6$, in order for the ejecta to be slowed down mainly by the dense clouds rather than by the low density medium that they are embedded in.

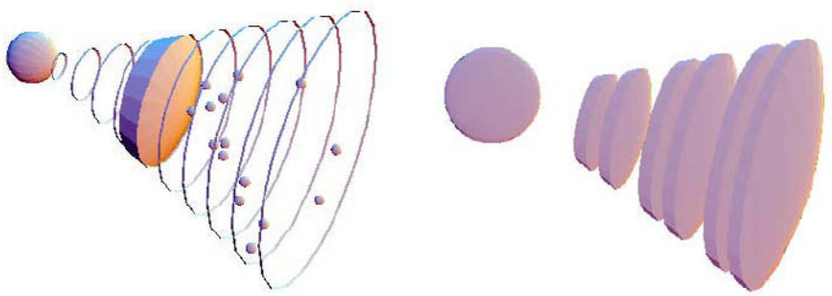

Fig. 9. *Producing variability by external shocks (left) or internal shocks (right). In the external shocks scenario, the variability is produced by irregularities in the surrounding. If the surrounding consists of a low density medium that contains high density clouds, then whenever the shell hits one of the clouds a peak in the emission is produced. The number of clouds, within the observable cone (of angular size* $1/\gamma$ *due to relativistic beaming) should therefore roughly be the number of observed peaks. The source itself, in this model, needs to produce only a single shell in a single (simple) explosion. However, the external shocks scenario has low efficiency, due to the small total surface area of the clouds when compared to the area of the shell. In the internal-shocks case, the temporal structure arises from the source, i.e. the source produces a more complex explosion. There is no efficiency problem, provided that the relative Lorentz factor between shells is large.*

Internal shocks do not suffer from these problems. Detailed calculations show that the observed temporal structure from internal shocks, closely follows the operation of the inner engine that generated the shells [57]. In this scenario, the source must be variable on time scales shorter than a second and last for as long as 100 seconds, just as the bursts themselves.

The efficiency of internal shocks is largely determined by the ratio of Lorentz factors between different shells which are colliding with each other. The larger the ratio, the larger the efficiency. A simple scenario that demonstrates this is the case of two equal mass shells with Lorentz factor $\gamma_1 \gg \gamma_2 \gg 1$. Conservation of energy and momentum in a collision between the shells leads to a Lorentz factor

which is the geometric mean of the initial ones $\sqrt{\gamma_1\gamma_2}$. Therefore, the energy left in the system as non thermal is a small fraction $\sqrt{\gamma_2/\gamma_1}$ of the initial energy. Beloboradov [58] has argued that if large Lorentz factor ratios are allowed, the internal shock efficiency is only limited by the fraction of energy in the shock given to the radiating electrons. Kobayashi and Sari [59] have then shown that multiple collisions between shocks may result in 'ultra efficient' internal shocks, in the sense that even more than the fraction of energy given to electrons can be radiated away.

The mechanism by which the thermal energy produced by internal shocks is converted to radiation is almost certainly synchrotron and inverse Compton, since these are the dominant radiation mechanisms at the low densities involved. While both mechanisms probably take place, it is actually not very clear which of the two produces the observed radiation. Synchrotron emission is for several reasons preferred [60,61] and inverse Compton probably produces a higher energy component.

4 The Afterglow: Theory

After the internal shocks produced the GRB, the shell interacts with the surrounding medium and decelerates. Again it emits radiation by synchrotron and inverse Compton. As the flow decelerates, the emission shifts to lower and lower frequencies. This emission, the afterglow, may last on detectable levels for years after the GRB event!

Afterglow was predicted well before it was observed [62,63,64,65]. The afterglow theory is relatively simple. It deals with the emission on timescales much longer than that of the GRB. The details of the complex initial conditions are therefore forgotten and the condition of the GRB remnant can be described by a self similar solution with a small number of parameters, such as the total energy and the external density. It is assumed that the electrons are accelerated by the shock into a power-law distribution (index p) of electron Lorentz factors $N(\gamma_e) \propto \gamma_e^{-p}$ for $\gamma_e > \gamma_m$. The lower cutoff, γ_m, of this distribution is set by the assumption that the electrons acquire a fixed fraction, ϵ_e, of the thermal energy (assumption of equipartition). It is also assumed that a considerable magnetic field is built behind the shock, which is again characterized by a certain fraction ϵ_B of equipartition. The energy density behind a relativistic shock is given by $4\gamma^2 n_1 m_p c^2$, where n_1 is the proton density ahead of the shock in units of cm^{-3}, γ is the Lorentz factor of the fluid behind the shock, and m_p is the proton mass. These equipartition assumptions then result in

$$\gamma_m = \frac{p-2}{p-1}\frac{m_p}{m_e}\epsilon_e\gamma \cong 630\epsilon_e\gamma \tag{2}$$

$$B = 0.4\sqrt{\epsilon_B n_1}\gamma \text{ Gauss}, \tag{3}$$

where B is the magnetic field, and m_e is the electron mass. The relativistic electrons then emit synchrotron radiation which produces the observed afterglow.

The broad band spectrum of such afterglow emission was given by Sari, Piran & Narayan [66].

The afterglow synchrotron spectrum can be fully described by the electron energy index p, the peak flux F_m and three characteristic frequencies (ν_m, ν_c, ν_a):

(I) ν_m is the synchrotron frequency of the minimal energy electron, with Lorentz factor γ_m. From synchrotron theory $\nu_m \cong (eB/2\pi m_e c)\gamma_m^2$ in the local frame of the fluid; here e is the electron charge. Transforming this to the observer frame (blue shifted by the Lorentz factor and redshifted by a factor of $[1+z]$) and using equations 2 and 3 we obtain

$$\nu_m = 1.4 \times 10^{13}\mathrm{Hz}\,(1+z)^{-1}\left(\frac{\epsilon_e}{0.1}\right)^2\left(\frac{\epsilon_B}{0.1}\right)^{1/2}(\frac{\gamma}{10})^4 n_1^{1/2}. \quad (4)$$

(II) The cooling time of an electron is inversely proportional to its Lorentz factor γ_e. Therefore, electrons with a Lorentz factor higher than a critical Lorentz factor $\gamma_e > \gamma_c$ cool on the dynamical timescale of the system. This characteristic Lorentz factor is given by the condition $\sigma_T c\gamma^2\gamma_c^2 B^2 t\gamma/6\pi(1+z) = \gamma_c m_e c^2$, and corresponds to the 'cooling frequency'

$$\nu_c = 1.2 \times 10^{13}\mathrm{Hz}\,(1+z)\left(\frac{\epsilon_B}{0.1}\right)^{-3/2}\left(\frac{\gamma}{10}\right)^{-4} n_1^{-3/2} t_{days}^{-2}, \quad (5)$$

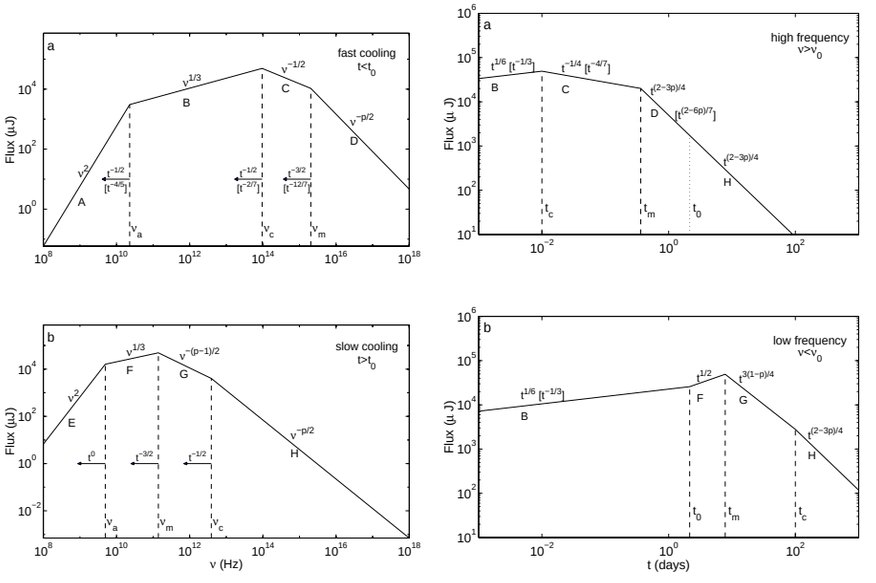

Fig. 10. *Theoretical spectra (left) and light curves (right) of synchrotron emission from a powerlaw distribution of electrons for the case of a constant density ambient medium and a spherical explosion. For most cases $p = 2.2 - 2.5$ fits the observed spectra and lightcurves well.*

where t_{days} is the observer time in days. Here we had also to take into account that time is redshifted.

(III) Below some critical frequency ν_a the flux is self absorbed and is given by the Rayleigh-Jeans portion of a black body spectrum[2]. The self-absorption frequency is given by

$$\nu_{sa} = 93\,\mathrm{GHz}\,(1+z)^{-13/5}\left(\frac{\epsilon_B}{0.1}\right)^{6/5}\left(\frac{\gamma}{10}\right)^{28/5} n_1^{9/5} t_{days}^{8/5}, \tag{6}$$

if $\nu_c < \nu_m$, and by

$$\nu_{sa} = 87\,\mathrm{GHz}\,(1+z)^{-8/5}\left(\frac{\epsilon_e}{0.1}\right)^{-1}\left(\frac{\epsilon_B}{0.1}\right)^{1/5}\left(\frac{\gamma}{10}\right)^{8/5} n_1^{4/5} t_{days}^{3/5}, \tag{7}$$

if $\nu_c > \nu_m$.

(IV) The normalization of the spectrum is given by the total number of radiating electrons $4\pi R^3 n_1/3$ times the peak flux from a single electron, resulting in

$$F_m = 220\,\mathrm{mJy}(1+z)^{-2} d_{L,28}^{-2}\left(\frac{\epsilon_B}{0.1}\right)^{1/2}\left(\frac{\gamma}{10}\right)^{8} n_1^{3/2} t_{days}^{3}, \tag{8}$$

where $d_{L,28}$ is the luminosity distance in units of 10^{28}cm.

The broad band spectrum of the well studied GRB 970508 [68] is in very good agreement with the theoretical picture. Note that the derivation above is quite general. It does depend neither on the surrounding density profile nor on the geometry of the event. Both these effects are hidden in the evolution of the fluid Lorentz factor γ, and the particle density n_1 as a function of time.

The evolution of this spectrum as a function of time depends on the hydrodynamics. The simplest model, which describes the observations in some cases quite well, is the adiabatic model with a constant density surrounding medium. The rest mass collected by the shock at radius R is about $R^3\rho$, where ρ is the mass density. On average, the particles move with a Lorentz factor of γ^2 in the observer frame (one factor of γ is the bulk motion and the other is the random thermal motion). Therefore, the total energy is given by $E \propto \gamma^2 R^3 \rho c^2$. Assuming that the radiated energy is negligible compared to the energy of the flow, we obtain that $\gamma \propto R^{-3/2}$ or in terms of the observer time, $t = R/\gamma^2 c$, we get $\gamma \propto t^{-3/8}$.

$$\nu_m = 6\times 10^{15}\,\mathrm{Hz}\;(1+z)^{1/2} E_{52}^{1/2} \epsilon_e^2 \epsilon_B^{1/2} t_{\mathrm{days}}^{-3/2}$$

$$\nu_c = 9\times 10^{12}\,\mathrm{Hz}\;(1+z)^{-1/2} \epsilon_B^{-3/2} n_1^{-1} E_{52}^{-1/2} t_{\mathrm{days}}^{-1/2}$$

$$\nu_{sa} = 2\times 10^{9}\,\mathrm{Hz}\;(1+z)^{-1} \epsilon_e^{-1} \epsilon_B^{1/5} n_1^{3/5} E_{52}^{1/5}$$

[2] Granot, Piran & Sari [67] have found that if $\nu_c < \nu_m$, then the self absorption frequency actually splits into two: ν_{ac} and ν_{sa}, where an optical depth of unity is produced by non-cooled electrons and all electrons, respectively. In between these two frequencies the spectral slope is $\nu^{11/8}$.

$$F_m = 20\,\mathrm{mJy}\ (1+z)\epsilon_B^{1/2} n_1^{1/2} E_{52} d_{L28}^{-2}$$

If, on the other hand, the density drops as R^{-2} (as is expected if the surrounding is a wind produced earlier by the progenitor of the burst) we get $\gamma \sim t^{-1/4}$. Choosing the parameter A_* to define the normalization of the density as $\rho R^2 = A_* 5 \times 10^{11} A_*$ gr/cm results in

$$\nu_m = 1.7 \times 10^{14}\,\mathrm{Hz}\ (1+z)^{1/2} E_{52}^{1/2} \epsilon_e^2 \epsilon_B^{1/2} t_{\mathrm{days}}^{-3/2}$$

$$\nu_c = 7 \times 10^{11}\,\mathrm{Hz}\ (1+z)^{-3/2} \epsilon_B^{-3/2} A_*^{-2} E_{52}^{1/2} t_{\mathrm{days}}^{1/2}$$

$$\nu_{sa} = 1.5 \times 10^{10}\,\mathrm{Hz}\ (1+z)^{-2/5} \epsilon_e^{-1} \epsilon_B^{1/5} A_*^{6/5} E_{52}^{-2/5} t_{\mathrm{days}}^{-3/5}$$

$$F_m = 180\,\mathrm{mJy}\ (1+z)^{3/2} \epsilon_B^{1/2} A_* E_{52}^{1/2} t_{\mathrm{days}}^{-1/2} d_{L28}^{-2}$$

These simple scalings, for the case of a constant density ambient medium, lead to the spectral evolution as given in Figure 10. The derivations above use a very simple description of the flow. It represents the fluid as if it has a single magnetic field strength and a single Lorentz factor γ and all of the material is moving directly towards the observer. Also, a very approximate description of the synchrotron emission was used. In reality, of course, the situation is more complicated. There are two effects that must be taken into account. The most dramatic one is the fact that matter slightly off the line of sight does not move directly towards the observer [69,123,71]. The amount of Lorentz boost from that matter is reduced. Secondly, fluid elements at different distances from the shock have somewhat different Lorentz factors, magnetic fields and electron energies. These variations can be estimated using the self-similar solution of Blandford and McKee [72]. The outcome of these more detailed calculations are the same scaling laws, but with a more accurate coefficient for the break frequencies as well as an estimate of the shape of the spectrum around each break frequency [73,74,75]. The equations given above already take these effects into account, and the coefficients given are accurate for $p = 2.2$.

The above scalings assumed adiabatic evolution. At first sight one may think that if the fraction of energy given to the electrons, ϵ_e, is less than unity, then perhaps only a small fraction of the energy can be radiated away. However, the same fireball energy is given again and again to newly shocked electrons. Each time, a fraction ϵ_e can be radiated away, and the overall effect can be large, much above the fraction ϵ_e. Energy losses during the cooling phase can be taken into account [71,76] using $dE/dR = -(16\pi/3)R^2 \epsilon_e \gamma^2 m_p c^2 n$. This results in $E = E_0 \times (t/t_0)^{-17\epsilon_e/12}$ for a constant density environment and $E = E_0 \times (t/t_0)^{-3\epsilon_e/2}$ for a wind environment. These effects are not taken into account in many models but may actually have a significant impact if ϵ_e is not too far below unity. In the case of GRB 000926, energy losses appear to have reduced the energy of the system by a factor of 5 [77].

Given the above hydrodynamic evolution, one can construct light curves at any given frequency. These will also consist of power laws, changing from one power law to the other once the break frequencies pass through the observed

band. These predicted power law lightcurves and spectra are in fair agreement with afterglow observations (see Sect. 5).

We have so far considered synchrotron radiation only. Since the optical depth of the system is small, most of the synchrotron photons emitted can be observed. Still, inverse Compton can affect the system in two ways. First, it may add an observable high-energy component. This requires a moderately high density. Second, it may provide an important cooling mechanism, and alter the synchrotron spectrum by its effect on ν_c. The ratio of the inverse Compton (IC) to synchrotron luminosity (a measure of their relative importance for cooling) can be computed very generally [60], in a way that does not deal with the details of the spectrum, but depends only on the underlying physical properties of the expanding shock wave. We generalize the derivation given by [60] to describe both fast and slow cooling regimes by introducing a parameter η, equal to the fraction of the electron energy that was radiated away (via both synchrotron and IC emission) [78]. Then the ratio of luminosities, in the limit of single scattering, is given by

$$x \equiv \frac{L_{IC}}{L_{syn}} = \frac{U_{rad}}{U_B} = \frac{U_{syn}}{U_B} = \frac{\eta U_e/(1+x)}{U_B} = \frac{\eta \epsilon_e}{\epsilon_B (1+x)}, \tag{9}$$

where U_{syn}, U_B and U_e are the energy density of synchrotron radiation, magnetic field and relativistic electrons, respectively. Note that in general $U_{syn} = \eta\beta U_e/(1+x)$, where β is the velocity of material behind the shock front (in the frame of the shock); however, for a relativistic shock $\beta \cong 1$. The importance of inverse Compton therefore diminishes quickly when the fireball becomes non relativistic.

Solving Eq. (9) for x we obtain

$$x = \frac{-1 + \sqrt{1 + 4\frac{\eta\epsilon_e}{\epsilon_B}}}{2}. \tag{10}$$

This solution has two interesting limits:

$$x = \begin{cases} \frac{\eta\epsilon_e}{\epsilon_B}, & \text{if} \quad \frac{\eta\epsilon_e}{\epsilon_B} \ll 1, \\ \left(\frac{\eta\epsilon_e}{\epsilon_B}\right)^{1/2}, & \text{if} \quad \frac{\eta\epsilon_e}{\epsilon_B} \gg 1. \end{cases} \tag{11}$$

Modeling afterglow data often suggests that $\epsilon_e \gg \epsilon_B$ and therefore inverse Compton may be of importance.

5 The Afterglow Revolution

Motivated by the prediction of a late-time softer radiation (the afterglow), several groups executed rapid radio follow-up observations of GRB error boxes. Detection of a radio afterglow seemed most promising. Not only does the large field of view match well with the large error boxes (several degrees) that were then available on short time scales (within a day), but maximum light was also expected to

occur later at longer wavelengths. The best (pre-BeppoSAX era) limits on such afterglow radio emission were obtained for GRB 940301. This GRB triggered an extensive multi-wavelength campaign with ground based optical and radio observatories from the BATSE/COMPTEL/NMSU Rapid Response Network (RRN; [79]). No obvious candidate radio counterparts were found [80,81,82].

5.1 The First Identifications

The breakthrough came in early 1997, when the Wide-Field Cameras (WFCs; [83]; see Fig. 11) onboard the Italian-Dutch satellite BeppoSAX [84] (see Fig. 12) obtained their first quickly available (within hours) accurate positions of GRBs (several arcminutes). This allowed rapid follow-up observations which led to the discoveries of X-ray [85], optical [86], millimeter [87] and radio [88] counterparts of GRBs. These observations quickly settled the distance controversy. The first transient optical counterpart, of GRB 970228, is in a faint galaxy with $\sim 0.8''$ diameter [89]. And, detection of absorption features in the OT's spectrum of GRB 970508 [90] established that this event was at a redshift greater than $z = 0.835$. GRBs come from 'cosmological' distances and are thus extremely powerful events. They are by far the most luminous photon sources in the Universe, with (isotropic) peak luminosities in γ rays up to 10^{52} erg/s, and total energy budgets up to several 10^{53-54} erg [91,92] (but see Sect. 6 and 7 for a discussion on collimated outflow, which reduces the inferred total energy). Within the first day, the optical emission is usually brighter than 20th magnitude (some 10 mag brighter [absolute] than the brightest supernovae) and therefore small telescopes can play an important role in measuring the lightcurve. Nowadays, a

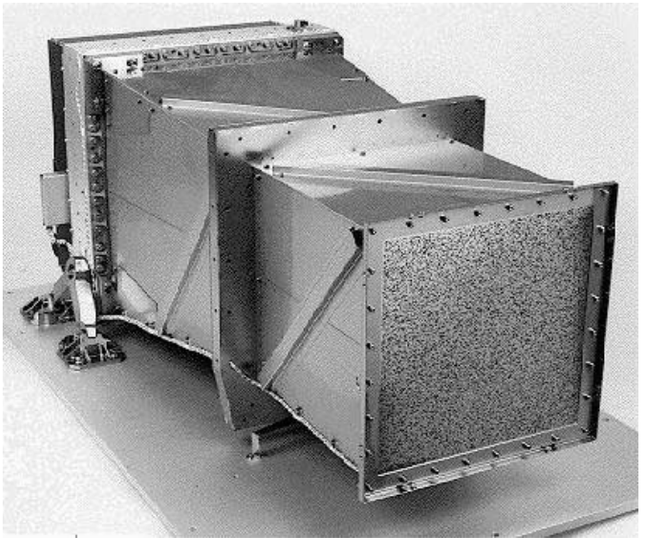

Fig. 11. *BeppoSAX Wide Field Camera (Courtesy SRON)*

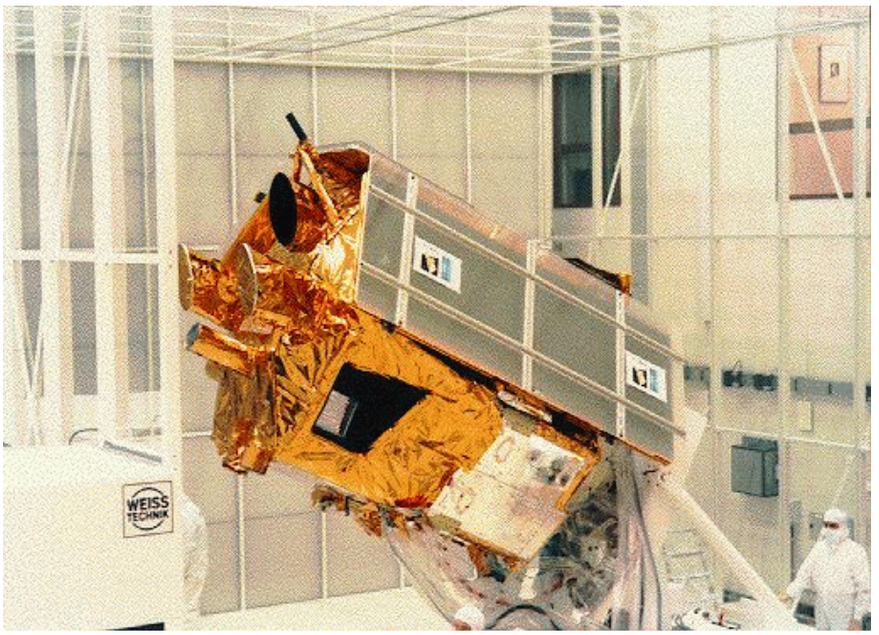

Fig. 12. *The Italian-Dutch BeppoSAX X-ray satellite. Visible are the Narrow Field Instruments (to the left) and, underneath a solar panel, one of the Wide Field Cameras.*

large worldwide collaboration is observing these events and the data are submitted to a Global-Coordinate-Network in real time, allowing other observatories to react rapidly.

5.2 Confirmation of the Relativistic Blast-Wave Model

A stringent test of the relativistic blast-wave model came with the discovery of X-ray [85] and optical afterglow following GRB 970228 [86,93]. The X-ray and optical afterglows of GRB 970228, show a power-law temporal decay; this is a trend observed in all subsequent X-ray and optical afterglows, with power-law exponents in the range 1 to 2.

Let us first concentrate on the forward shock and assume slow cooling (the bulk of the electrons do not radiate a significant fraction of their own energy and the evolution is adiabatic); this appears applicable to some observed GRB afterglows at late times ($t > 1$ hr). The simplest assumption is that of spherical symmetry and a constant ambient density. As both the afterglow's spectrum and the temporal evolution of the break frequencies ν_a, ν_m, ν_c are, in the relativistic blast-wave model, power laws (see Sect. 4), the evolution of the flux is also a power law in time. For example, for $\nu_m \leq \nu \leq \nu_c$, the decay of the flux is $F_\nu \propto t_{\rm obs}^{-3(p-1)/4}$, and the power law spectral slope α relates to the spectral slope β

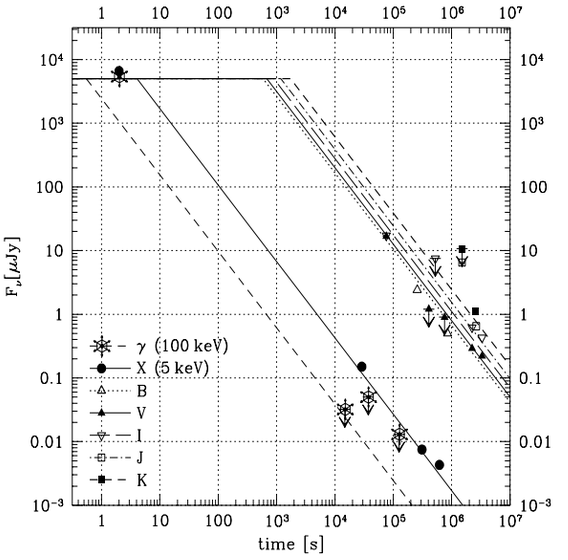

Fig. 13. *The light curves of GRB 970228 from gamma rays to near infrared (from [94]). To first order the light curves are power laws and the offsets between them satisfy the expectations from the model.*

as $\alpha = -3/2\beta$. Several authors [94,95,96] showed that to first order this model describes the X-ray and optical afterglow of GRB 970228 very well (see Fig. 13).

GRB 970508 was the first GRB with a radio counterpart [88]. The radio light curves (8.5 and 4.9 GHz) show large variations on time scales of less than a day, but these damp out after one month. This finds a viable explanation in interstellar scintillation (irregular plasma refraction by the interstellar medium between the source and the observer). The damping of the fluctuations can then be understood as the effect of source expansion on the diffractive interstellar scintillation. Thus a source size of roughly 10^{17} cm was derived (at 3 weeks), corresponding to a mildly relativistic expansion of the shell [88].

GRB 970508 remains one of the best observed afterglows: the radio afterglow was visible at least 368 days (and at 2.5 sigma on day 408.6 [97]), and the optical afterglow up to $\sim$ 450 days (e.g. [98,99,100]). In addition millimeter [87], infrared and X-ray [101] counterparts were detected, and it is the first GRB for which a spectral transition in the optical/near IR range was found [68,99]; this transition is interpreted as the effect of the passage of the cooling frequency through the optical/near IR passbands. These multiwavelength observations allowed the reconstruction of the broad radio to X-ray spectrum for this GRB [68] (see Fig. 14). Galama et. al. [68] found that the 'standard' model provides a successful and consistent description of the afterglow observations over nine decades in frequency, ranging in time from the event until several months later. The synchrotron afterglow spectrum of this GRB allows measurement of the

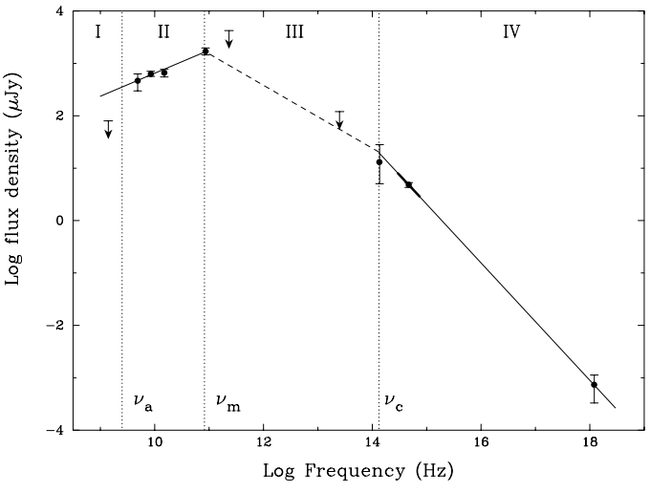

Fig. 14. *The X-ray to radio spectrum of GRB 970508 on May 21.0 UT (12.1 days after the event). The location of the break frequencies ν_a, ν_m and ν_c, inferred from transitions in the light curves and from spectra of the afterglow, are indicated (from [68]).*

electron energy spectrum p, the three break frequencies (ν_a, ν_m and ν_c), and the flux at the peak, F_m. For GRB 970508 the redshift, z, is also known, and all blast wave parameters could be deduced: the total energy (per unit solid angle) E = $3.5{\times}10^{52}$ erg, the ambient (nucleon) density $n_1 = 0.030$, the fraction of the energy in electrons $\epsilon_e = 0.12$ and that of the magnetic field ϵ_B= 0.089 [102]. The numbers themselves are uncertain by an order of magnitude (see e.g., [75]), but the result shows that the 'standard' model fits the expectations very well.

Following these first attempts at modeling the broad-band afterglow more detailed modeling efforts have been made. For example, Panaitescu and Kumar [103] have modeled a sample of GRBs with relativistic jets (see Sect. 6 and 7 for a detailed discussion on jets) and find: typical energies of $10^{50} - 10^{51}$ erg, ambient densities ranging from $10^{-3} - 10$ cm^{-3}, beaming angles ranging between $1^\circ - 4^\circ$, and that a wind-like ambient medium can in some cases be ruled out (see Sect. 10.6 for a detailed discussion on wind-like media) but in others it can not. GRB 000301C was modeled with a hard electron-energy distribution ([105]; $p = 1.5$) (but see [106]) and also GRB 010222 requires a hard electron-energy distribution [107]. And, evidence has been presented for an inverse Compton emission component in the afterglow of GRB 000926 [104].

The highly relativistic nature of the GRB source [108] can once more be seen in the extreme brightness temperature of the GRB 990123 optical flash ([45] $T_b \gtrsim 10^{17}$ K; see Sect 9) which by far exceeds the Compton limit of 10^{12} K. In this case the optical signal from GRB 990123 was some 18 mag brighter (absolute) than the brightest supernovae. The extreme brightness can be explained by emission from the reverse shock (see Sect. 9).

5.3 Short-Duration GRBs

All knowledge derived from afterglow detections to date (distance scale, progenitor distribution, etc.) applies only to long-duration ($t \gtrsim 2$ s) GRBs, as short-duration GRBs have never been precisely localized quickly enough to allow follow-up observations. There is good evidence that the short GRBs comprise a distinct population in terms of their γ-ray properties, and it is therefore likely that GRBs derive from multiple progenitor types. It is possible that short GRBs may result from mergers, and, residing in low-density regions, have very faint afterglows [109].

6 Collimated Outflow (Jets): Theory

The hydrodynamic evolution described in Sect. 4, assumed spherical symmetry. However, many astrophysical phenomena, especially those involving extreme energetics are not spherical but in the form of jets. As we will see, this is most probably the case also for GRBs.

Jets have been discussed extensively in the context of GRBs. First, the similarity between some of the observed features of blazars and AGNs led to the speculation that jets also appear in GRBs [110]. Second, the regions emitting the GRBs as well as the afterglow must be moving relativistically. The emitted radiation is strongly beamed, and we can observe only a region with an opening angle $1/\gamma$ off the line of sight. Emission outside of this very narrow cone is not observed. These considerations have lead to numerous speculations on the existence of jets and to attempts to search for the observational signature of jets both during the GRB phase [111] and in the context of the afterglow [112,113,114]. Finally, jets appear naturally in the context of several leading scenarios for the 'inner engine' (see Sect. 10).

We begin by clarifying some of the confusing terminology. There are two distinct, but related, effects. The first, *'jets'*, describes scenarios in which the relativistic flow emitted from the source is not isotropic but collimated towards a finite solid angle. The term jet refers to the geometrical shape of the relativistic flow emitted from the inner engine. The second effect is that of *'relativistic beaming'*. The radiation from any object that is radiating isotropically in its own rest frame, but moving with a large Lorentz factor γ in the observer frame, is beamed into a small angle $1/\gamma$ around its direction of motion. This is an effect of special relativity. It has nothing to do with the ejecta's geometry (spherical or jet) but only with the fact that the ejecta is moving relativistically. The effect of relativistic beaming allows an observer to see only a small angular extent, of size $1/\gamma$ centered around the line of sight. Since we know the flow is ultra-relativistic (initially $\gamma > 100$), there is no question that the relativistic beaming effect is always relevant for GRBs. The question we are interested in is that of the existence of 'jets'.

The idealized description of a jet is a flow that occupies only a conical volume with half opening angle θ_0. In fact, the relativistic dynamics is such that

the width of the material in the direction of its propagation is much smaller than its distance from the source by a factor of $1/\gamma^2$. The flow, therefore, does not fill the whole cone. Instead it occupies only a thin disk at its base, looking more like a flying pancake [116] (see Figure 9). If the 'inner engine' emits two such jets in opposite directions then the total solid angle towards which the flow is emitted is $\Omega = 2\pi\theta_0^2$. Whether the relativistic flow is in the form of a jet or a sphere has three important implications.

The Total Emitted Energy. Optical observations of afterglows enabled redshift determination, and therefore a reasonably accurate estimate of the distance, D, to these events (the uncertainty is now in the cosmological parameters of the Universe). The so called 'isotropic energy' can then be inferred from the fluence F (the total observed energy per unit area at earth) as $E_{iso} = 4\pi D^2 F$ (taking cosmological corrections into account, $D = d_L/\sqrt{1+z}$ where d_L is the luminosity distance and z is the redshift). The numbers obtained in this way range from 10^{51} erg to 10^{54} erg with the record of 3×10^{54} erg held by the famous GRB 990123. These huge numbers approach the equivalent energy of a solar mass, all emitted in a few tens of seconds!

These calculations assumed that the source emitted the same amount of energy towards all directions. If instead the emission is confined to some solid angle Ω then the true energy is $E = \Omega D^2 F$. As we show later Ω is very weakly constrained by the GRB itself and can be as low as 10^{-6}. If so the true energy in each burst $E \ll E_{iso}$. We will show later that interpretation of the multi-wavelength afterglow lightcurves indeed indicates that some bursts are jets with solid angles considerably less than 4π. The isotropic energy estimates may be fooling us by a few orders of magnitudes! Clearly this is of fundamental importance when considering models for the sources of GRBs.

The Event Rate. In its glory days, BATSE detected about one burst per day. With the help of several redshift measurements, or alternatively, with the use of the cumulative brightness distribution (the Log N/ Log S curve), this translates to about 10^{-7} bursts per year per galaxy or 0.5 bursts/Gpc^{-3}/year [117,118]. However, if the emission is collimated to $\Omega \ll 4\pi$ then we do not see most of the events. The true event rate is then larger than that measured by BATSE by a factor of $4\pi/\Omega$. Again this is of fundamental importance. Clearly, the corrected GRB event rate must not exceed that of compact binary mergers or the birth rate of massive stars if these are to produce the majority of the observed GRBs (see Sect. 10).

The Physical Ejection Mechanism. Different physical models are needed to explain collimated and isotropic emission. For example, in the collapsar model (e.g. [115]), relativistic ejecta, that is believed to create the GRB, is produced only around the rotation axis of the collapsing star with half opening angle of about $\theta_0 \cong 0.1$. Such models would have difficulties explaining isotropic bursts as well as very narrow jets.

6.1 The Jet-Break

As the afterglow evolves, γ decreases and it will eventually fall below the initial inverse opening angle of the jet. The observer will notice that some of the sphere is missing from the fact that less radiation is observed. This effect alone, will produce a significant break, steepening the lightcurve decay by a factor of $\gamma^2 \propto t^{-3/4}$ even if the dynamics of each fluid element have not changed. The transition should occur at the time t_{jet} when $1/\gamma \cong \theta_0$. Observing this time can therefore provide an estimate of the jet's opening angle according to

$$t_{\rm jet} \approx 6.2{\rm hr}(1+z)(E_{52}/n_1)^{1/3}(\theta_0/0.1)^{8/3} \tag{12}$$

Additionally, Rhoads [113] has shown that at about the same time (see however [124,120,121]), the jet will begin to spread laterally so that its opening angle $\theta(\grave{t}) \sim 1/\gamma$. The ejecta now encounters more surrounding matter and decelerates faster than in the spherical case. The Lorentz factor then decays exponentially with the radius and as $\gamma \propto t^{-1/2}$ with observed time. Taking this into account, the observed break is even more significant. The slow cooling spectrum given in Figure 10 evolves with decreasing peak flux $F_m \propto t^{-1}$ and the break frequencies evolve as $\nu_m \propto t^{-2}$, $\nu_c \propto t^0$ and $\nu_a \propto t^{-1/5}$. This translates to a temporal decay in a given frequency as listed in Table 1.

Table 1. *The spectral index β and the temporal index α as function of p for a spherical and a jet-like evolution. Typical values are quoted using $p = 2.4$. The parameter free relation between α and β is given for each case (eliminating p). The difference in α between a jet and a sphere is always substantial at all frequencies.*

	spectral index β, $F_\nu \propto \nu^{-\beta}$	light curve index α, $F_\nu \propto t^{-\alpha}$	
		sphere	jet
$\nu < \nu_a$	$\beta = -2$	$\alpha = -1/2$	$\alpha = 0$
$\nu_a < \nu < \nu_m$	$\beta = -1/3$	$\alpha = -1/2$	$\alpha = 1/3$
$\nu_m < \nu < \nu_c$	$(p-1)/2 \cong 0.7$	$\alpha = 3(p-1)/4 \cong 1.05$ $\alpha = 3\beta/2$	$\alpha = p \cong 2.4$ $\alpha = 2\beta + 1$
$\nu > \nu_c$	$p/2 \cong 1.2$	$\alpha = (3p-2)/4 \cong 1.3$ $\alpha = 3\beta/2 - 1/2$	$\alpha = p \cong 2.4$ $\alpha = 2\beta$

The jet break is a hydrodynamic one. It should therefore appear at the same time at all frequencies - an achromatic break[3]. Though an achromatic break is considered to be a strong signature of a jet, one should keep in mind that any other hydrodynamic transition will also produce an achromatic break. To name a few: the transition from relativistic to non-relativistic dynamics, a jump in the ambient density or the supply of new energy from slower shells that catch

[3] Sari 1997 [122], argued that there may be about a factor of two difference in the effective transition time between the four different spectral regimes (e.g. below or above ν_m) due to the fact that the emission in these different regimes weighs differently contributions from various emission radii.

up with the decelerated flow. However, the breaks produced by the transition from a spherical like evolution (when $1/\gamma < \theta_0$) to a spreading jet have a well defined prediction for the change in the temporal decay indices. The amount of break depends on the spectral regime that is observed. It can be seen from Table 1 that the break is substantial ($\Delta\alpha > 0.5$ in all regimes) and should be easily identified.

7 Observational Evidence for Collimated Outflow (Jets)

The theory of jets evolution and of the resulting light curves has been worked out before evidence for jets was obtained. In fact, Rhoads [113], has used this theory to constrain the amount of collimation in GRB 970508, which did not show any significant steepening of the afterglow lightcurve. He concluded that the opening angle of a jet, if it exists, must be more than 30 degrees. We note that if the jet's opening angle is of order unity, the total energy may still be about an order of magnitude lower than the isotropic estimate. However, in this case the break will be 'hidden' as it will overlap the transition to non-relativistic dynamics. Based on late time radio data, it was suggested that this is the case for GRB 970508 [97].

The first claim for narrow jets in GRBs came from Sari, Piran and Halpern [123]. They noted that the observed decays in GRB afterglows that do not show a break are either of a shallow slope $F_\nu \propto t^{-1.2}$ or a very steep slope $F_\nu \propto t^{-2}$. They argued that the rapidly decaying bursts are those in which the ejecta was a narrow jet and the break in the light curve was before the first observations. Interestingly, evidence for jets are found when the inferred energy (without taking jets into account) is the largest. This implies that the jets account for a considerable fraction of the wide luminosity distribution seen in GRBs, and the true energy distribution is less wide than it seems to be.

The predicted light-curve transition (from a regular to a fast decay caused by a jet) has been observed in the optical afterglow of GRB 990123 [92,124,125]. However, no evidence for such an increase of the decay rate was found in near-infrared K-band observations [92]. A similar transition was better sampled in afterglow data of GRB 990510; optical observations of GRB 990510, show a clear steepening of the rate of decay of the light simultaneously in all optical bands between ~ 3 hours and several days [126,127] to roughly $\mathrm{F}_\nu(t) \propto t^{-2.2}$. Together with radio observations, which also reveal a transition, it is found that the transition is very much frequency-independent; this virtually excludes explanations in terms of the passage of the cooling frequency, but is what is expected in case of beaming [126]. Harrison et al.[126] derive a jet opening angle (from the jet-break time) of $\theta_0 \cong 0.08$, which for this burst would reduce the total energy in γ rays to $\sim 10^{51}$ erg.

Frail et al. [144] have recently determined the jet-break times for a sample of GRBs with known redshifts. From these, a wide range of jet-opening angles is inferred in GRBs: from 3° to more than 25°, with a strong concentration near 4°. This relatively narrow collimation implies that the observed GRB rate has

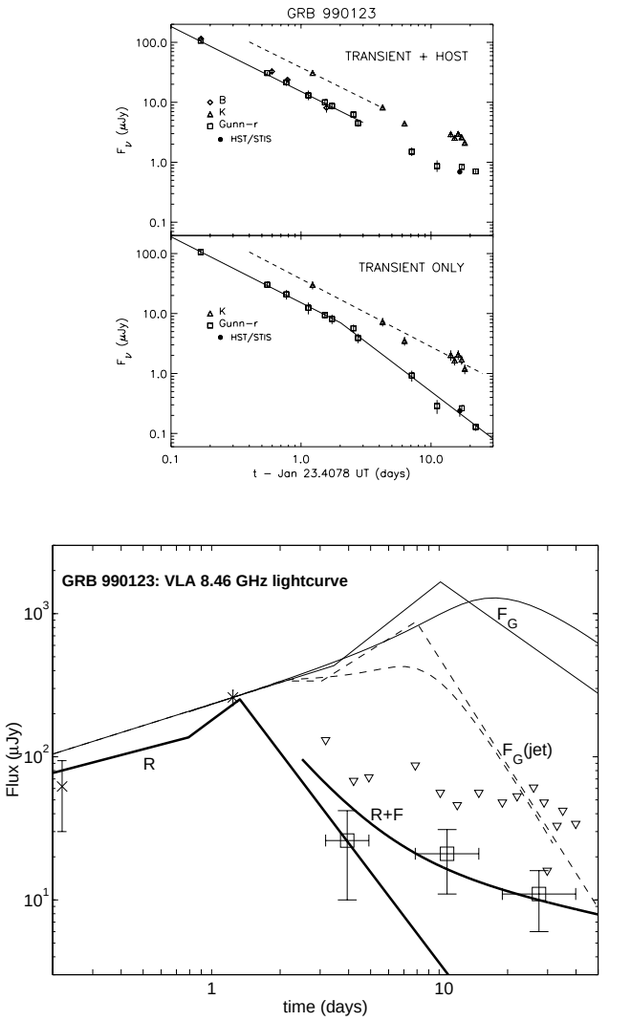

Fig. 15. *GRB 990123: Optical data (left) shows a slight break in the light curve at Gunn-r band. K band data shows no clear break, but the contribution of the host galaxy is less certain at K band. A radio 'flare' (right) is seen a day after the burst, and agrees with the theoretical scaling of the optical flash (heavy solid line marked R). In the jet interpretation, faint radio emission is only expected at late times (heavy solid line marked R+F). The theoretical expectations if the radio signal at day two were interpreted as resulting from the forward shock (independent of the optical flash) and in case jets are not taken into account (thin and dashed lines) will largely over-predict the late radio upper limits (marked by triangles) [143] (see however [108]).*

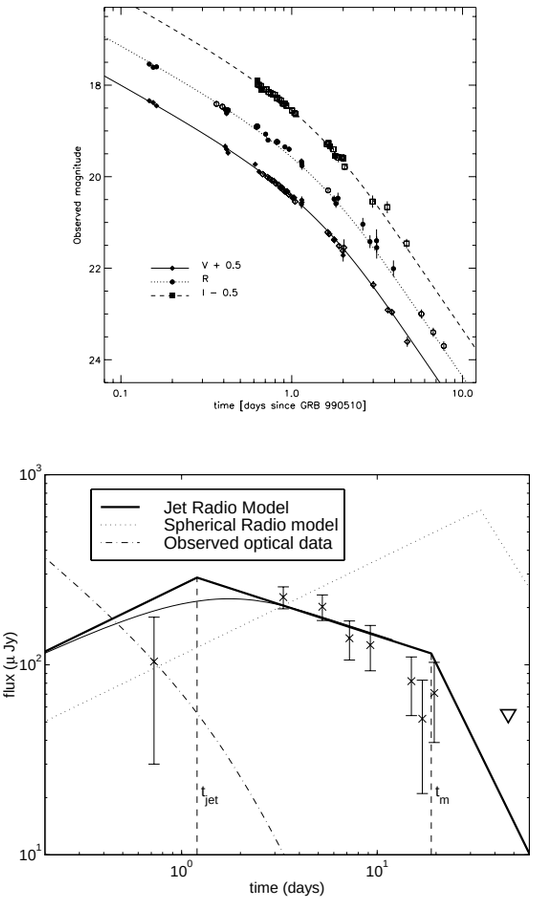

Fig. 16. *GRB 990510, the 'classical' case for a 'jet': an achromatic break in optical and radio at* $t_{\rm jet} = 1.2$ *days implying a jet-opening angle* $\theta_0 = 0.08$*. The temporal slope before and after the break agree well with the theory if* $p = 2.2$*. For this burst the isotropic gamma-ray energy* $E_{\rm iso} = 2.9 \times 10^{53}$ *erg but the 'true' total energy is only* $E = 10^{51}$ *erg. From [126].*

to be corrected for the fact that conical fireballs are visible to only a fraction of observers. Frail et al. find that the 'true' GRB rate is ~ 500 times larger than the observed GRB rate. Although the isotropic equivalent energies of GRBs range from about 5×10^{51} to 1.4×10^{54} erg, when one corrects the observed γ-ray energies for the geometry of the outflow, GRB energies appear narrowly clustered around 5×10^{50} ergs (see Fig. 17).

The central engines of GRBs thus produce approximately a similar amount of energy, and the broad range of fluence and luminosity observed for GRBs appears largely the result of a wide variation of opening angles. The reason for

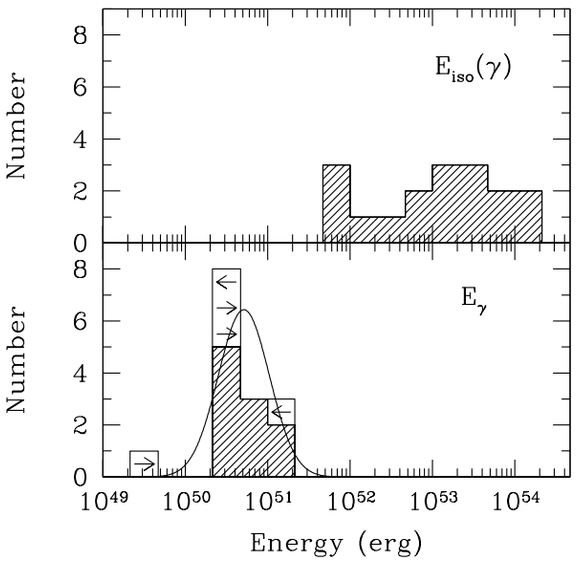

Fig. 17. *The distribution of the apparent isotropic γ-ray burst energy of GRBs with known redshifts (top) versus the geometry-corrected energy (bottom). While the isotropic energy E_{iso} spans three orders of magnitudes, the geometrically corrected energy, $E_\gamma = E_{iso}\theta^2/2$, is very narrowly distributed. This implies that the sources of GRBs produce roughly the same amount of energy, about 5×10^{50} erg, but that energy is distributed over a variety of angles resulting in a wide distribution of isotropic energies. From [144].*

why this range in angles exists is currently not understood. Gamma ray bursts have gone a long way in the past four years. It is interesting to note that before the redshift era, most models assumed that the events were standard candles with energies of about 10^{51} erg. As more and more redshifts are determined, the energy record increased steadily up to 10^{54} erg. The standard-candle hypothesis was abandoned. It is remarkable, that now, when more detailed understanding allows us to infer the beaming angles of these explosions, the true energy budget is back at $\sim 10^{51}$ erg, and the explosions are once again standard candles (though not in the same sense as before).

8 Polarization – A Promising Tool

An exciting possibility to further constrain the models and obtain a more direct proof of the geometrical picture of 'jets' is to measure linear polarization. Varying polarization at optical wavelengths has been observed in GRB afterglows at the level of a few percent [129,130].

High levels of linear polarization are usually the smoking gun of synchrotron radiation. The direction of the polarization is perpendicular to the magnetic field and can be as high as 70%. Gruzinov and Waxman [74] and Medvedev and Loeb [131] considered the emission from spherical ejecta which by symmetry

should produce no polarization on the average, except for fluctuations of order a few percent. Polarization is more natural if the ejecta is a 'jet' and the line of sight to the observer is within the jet but does not coincide with its axis. In this case, the spherical symmetry is broken [132,133,134], and the polarization produced by synchrotron radiation will not vanish. For simplicity, lets assume that the magnetic field behind the shock is directed along the shock's plane (the results hold more generally, as long as the magnetic field has a preferred direction). The synchrotron polarization from each part of the shock front, which is perpendicular to the magnetic field, is therefore directed radially.

As long as the relativistic beaming angle $1/\gamma$ is narrower than the physical size of the jet θ_0, one is able to see a full ring and therefore the radial polarization averages out (the first frame, with $\gamma\theta_0 = 4$ of the left plot in Figure 18). As the flow decelerates, the relativistic beaming angle $1/\gamma$ becomes comparable to θ_0 and only a part of the ring is visible; net polarization is then observed. Note that due to the radial direction of the polarization from each fluid element, the total polarization is maximal when a quarter ($\gamma\theta_0 = 2$ in Figure 18) or when three quarters ($\gamma\theta_0 = 1$ in Figure 18) of the ring are missing (or radiate less efficiently) and vanishes for a full and for half ring. The polarization, when more than half

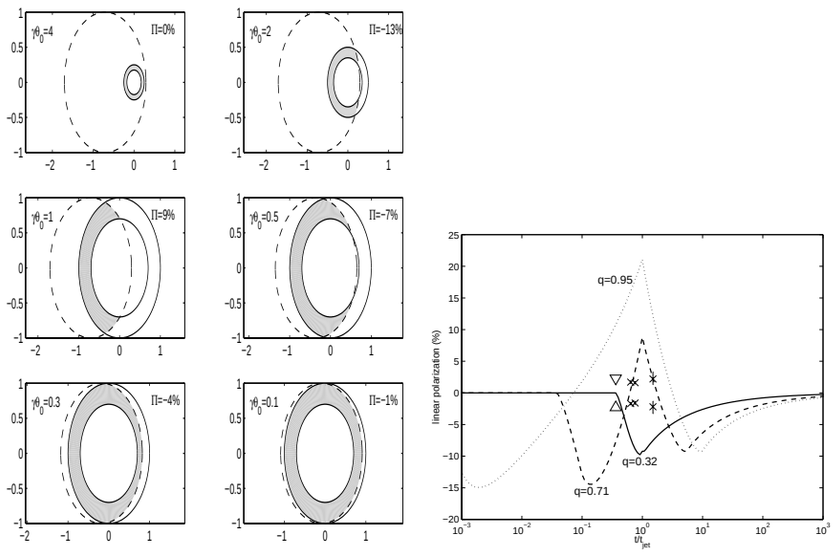

Fig. 18. *Left: Shape of the emitting region. The dashed line marks the physical extent of the jet, and solid lines give the viewable region $1/\gamma$. The observed radiation arises from the gray-shaded region. In each frame, the percentage of polarization is given at the top right and the initial size of the jet relative to $1/\gamma$ is given on the left. The frames are scaled so that the size of the jet is unity. Right: Observed and theoretical polarization lightcurves for three possible offsets of the observer relative to the jet axis Observational data for GRB 990510 is marked by crosses (x), assuming $t_{jet} = 1.2$ days. The upper limit for GRB 990123 is given by a triangle, assuming $t_{jet} = 2.1$ days.*

of the ring is missing, is perpendicular to the polarization direction when less than half of it is missing.

At late stages the jet expands sideways and since the offset of the observer from the physical center of the jet is constant, spherical symmetry is regained. The vanishing and re-occurrence of significant parts of the ring results in a unique prediction: there should be three peaks of polarization, with the polarization position angle during the central peak rotated by 90° with respect to the other two peaks. In case the observer is very close to the center, more than half of the ring is always observed, and therefore only a single direction of polarization is expected. A few possible polarization light curves are presented in Figure 18.

9 The Reverse Shock Emission: Theory and Observations

The previous sections discussed the theory and the observations of the 'late' afterglow, hours or more after the burst. During that time, most of the energy of the system was already given to the shocked surroundings, and it is that region that dominates the emission. However, during the first few tens of seconds, the evolution of the Lorentz factor as a function of time is not self similar. There are two shocks: a forward shock going into the surrounding medium and a reverse shock going into the expanding ejecta (see Figure 19). The hydrodynamic details were discussed in [135].

During the initial stages, the internal energy stored behind the shocked-surrounding matter and the energy of the shocked ejecta are comparable. However, the temperature of the shocked ejecta is much lower, typically by a factor of $\gamma \sim 10^2$. This results in an additional emission component with a typical frequency lower by a factor of $\gamma^2 \sim 10^4$, which, for typical parameters, is near the optical passband. Contrary to the 'standard' late afterglow, this emission is very sensitive to the initial Lorentz factor. Theoretical predictions for such a flash were given in detail by Sari & Piran [136,140] and were earlier suggested as a possibility by Mészáros & Rees [139].

One of the most exciting events in the field of afterglow studies, was the detection of bright (9th magnitude) optical emission simultaneous with GRB 990123 by the ROTSE team [45]. The ROTSE telescope obtained its first images only 22 seconds after the start of GRB 990123 (i.e. during the GRB), following a notification received from the BATSE aboard the Compton-satellite. The ROTSE observations show that the optical light curve peaked at $m_V \sim 9$ magnitudes some 60 seconds after the event [45]. After maximum a fast decay followed for at least 15 minutes. The late-time afterglow observations show a more gradual decline [108,92,124,125,137] (see Fig. 20).

The redshift $z = 1.6$, inferred from absorption features in the OT's spectrum, implies that the optical flash would have been as bright as the full moon had the GRB occurred in the nearby galaxy M31 (Andromeda). A different way to put this in perspective is that the flash was some 18 mag brighter (absolute) than the brightest supernovae. Galama et al. [108] have shown, that if one assumes

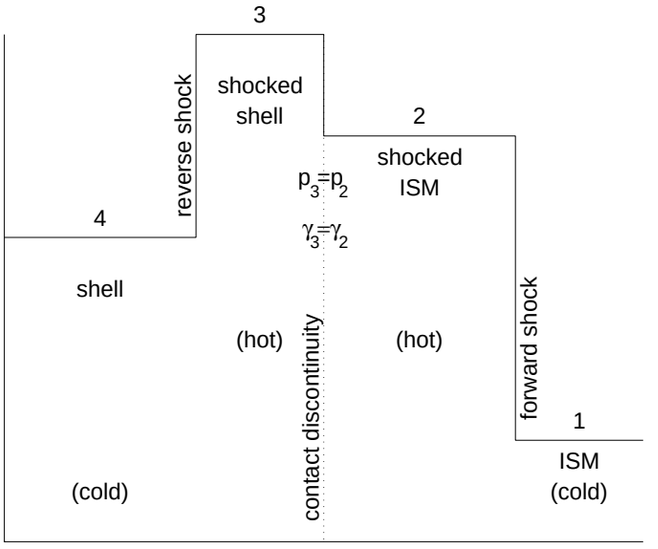

Fig. 19. *Schematic representation of the four zones that are present when a relativistic fireball starts to interact with the interstellar medium (ISM). The solid line indicates the density as a function of radius. The undisturbed ISM is zone 1 at large radius, and the unshocked shell material is zone 4 at small radii. The shocked zones, 2 and 3, are the result of the forward and reverse shocks and are separated by a contact discontinuity (based on Sari & Piran 1995).*

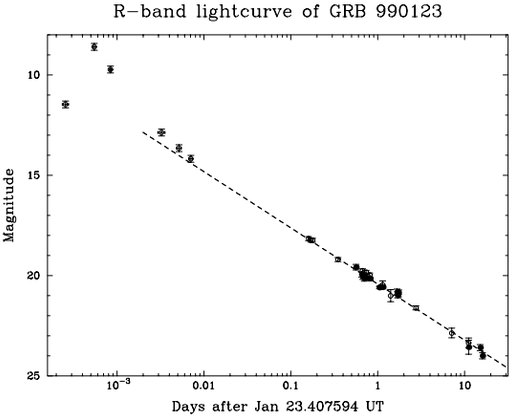

Fig. 20. *R-band light curve of the afterglow of GRB 990123. The ROTSE data show that the optical light curve peaked at* $m_V \sim 9$ *[45]. The dashed line indicates a power law fit to the light curve (for* $t > 0.1$ *days), which has exponent* -1.12 ± 0.03 *(from [108]).*

that the emission detected by ROTSE comes from a non-relativistic source of size ct, that then the observed brightness temperature $T_b \gtrsim 10^{17}$ K of the optical flash exceeds the Compton limit of 10^{12} K. This confirms the highly relativistic nature of the GRB source.

The observed optical properties of this event are well described by emission from the reverse shock that initially decelerates the ejecta, provided that the initial Lorentz factor is about 200 [137,141]. It takes tens of seconds for the reverse shock to sweep through the ejecta and produce the bright flash. Later, the shocked hot matter expands adiabatically and the emission quickly shifts to lower frequencies and considerably weakens.

The ROTSE observations show that the prompt optical and γ-ray light curves do not track each other [45]. In addition, detailed comparison of the prompt optical emission with the BATSE spectra of GRB 990123 (at three epochs for which both optical and gamma-ray information is available) shows that the ROTSE emission is not a simple extrapolation of the GRB spectrum to much lower energies [108,138].

If this interpretation is correct, GRB 990123 would be the first burst in which all three emitting regions have been seen: internal shocks causing the GRB, the reverse shock causing the prompt optical flash, and the forward shock causing the afterglow. The emissions thus arise from three different emitting regions, explaining the lack of correlation between the GRB, the prompt optical and the late-time optical emission [108] (but see [142]).

Another new ingredient that was found in GRB 990123 is a radio flare [143]. Contrary to all other afterglows, where the radio peaks around a few weeks and then decays slowly, this burst had a fast rising flare, peaking around a day and then decaying quickly. This can be interpreted as emission from the cooling ejecta that was earlier on heated by the reverse shock. Using the Blandford and McKee [72] self-similar solution to derive the evolution of the ejecta and its emission properties one finds that the typical frequency scales as $\nu_m^r \propto t^{-73/48}$ and the flux at that frequency scales as $F_m^r \propto t^{-47/48}$ [140] (see [59] for revised scalings when the temperature of the ejecta is non–relativistic). Therefore, within a day the emission from the adiabatically cooling ejecta that produced the 60s optical flash in GRB 990123 is expected to shift to radio frequencies [137]. Using the observed optical flash, and the above scalings, a good fit to the radio data is obtained. The optical flash and the radio flare may therefore be related.

Given the above interpretation of the reverse shock emission, it is important to ask whether GRB 990123 is an exception, or whether the phenomena of radio flares and optical flashes is more common. Radio flares appear to exist in other cases [128]. However, since early radio data is usually sparse, and these events did not have an early optical observation to find the associated optical flash, the interpretation in terms of emission from the reverse shock is less secure than in the case of GRB 990123. In the optical, from robotic optical experiments such as ROTSE and LOTIS, strong upper limits exists for several bursts. The upper limits show that the optical flash does not scale with the fluence of the event [145,146]. However, with reasonably small changes in the density or the

initial Lorentz factor, those events could have escaped detection [147]. Once future satellites, such as HETE-II and SWIFT, provide accurate positioning on timescales of seconds, strong constraints on the generality of optical flashes and radio flares will be possible to obtain.

Finally, we mention the possibility of a source operating on an afterglow timescale. In this case the reverse shock continues for a long time, newly shock-heated electrons are continuously created, and the emission in a significant part of the spectrum, especially in the sub-mm regime [148,149], will be dominated by the reverse shock. A similar situation will occur if the source has emitted a slower, but energetic shell, which later catches up with the decelerating ejecta. The relation between the reverse and forward shock spectrum in such a case was given in [149] and shown in Figure 21.

10 Progenitors

Observationally it is hard to distinguish between models of progenitors for GRBs. The GRB and the afterglow are produced when relativistic ejecta are slowed down, and no observable radiation emerges directly from the 'hidden engine' that powers the GRB. Thus, in spite of all discoveries the origin of GRBs has remained rather mysterious. Popular models for the origin of GRBs that (in principle) can provide the required energies, come in two classes: (i) compact object mergers such as the neutron star-neutron star (e.g., [150]) and neutron star-black hole mergers [52,151,152], and (ii) the core collapses of very massive stars (termed collapsars, 'failed' supernovae or hypernovae [153,154]).

The expectation is that compact object mergers give rise to a short-duration GRB (timescales of less than a second), while collapsars can produce long-duration GRBs (timescales much longer than a second). It is therefore popularly speculated that the short-duration GRBs are caused by merger systems, while the longer ones are due to the core collapses of very massive stars. Note that so far counterparts to short-duration GRBs have not been identified as short-duration GRBs have never been precisely localized quickly.

In the following we will summarize the observational predictions of different progenitor models and what we have learned from such observations. Throughout this entire section on progenitors please keep in mind that the nature of the evidence is suggestive rather than conclusive.

10.1 The Duration of the Event

In the previous sections we have seen that internal shocks imply that the source is variable on < 1 s timescales but lasts for tens of seconds in the case of long bursts. This rules out progenitor models that lead to simple single explosions, as the engine needs to persist for longer times. Both compact object mergers and collapsars can fulfill this requirement as the scenario for GRB production usually involves the formation of a disk that persists for sufficiently long time and feeds a newly born black hole.

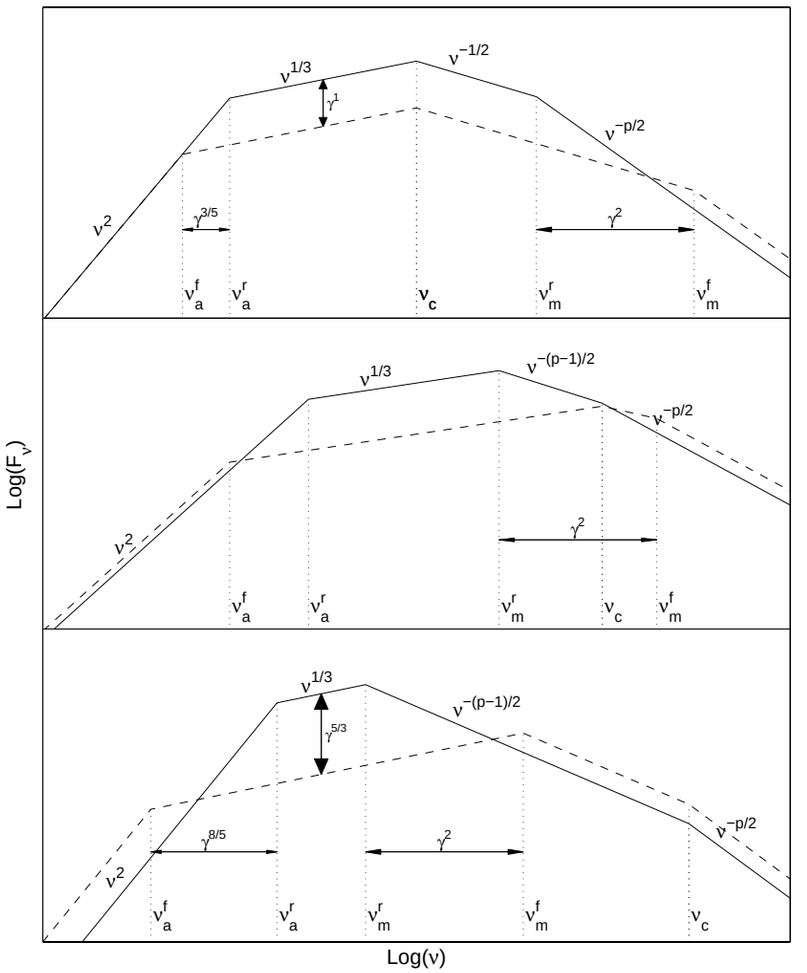

Fig. 21. *The combined reverse-forward shock spectra with synchrotron peaks at ν_m^r and ν_m^f, synchrotron self-absorption frequencies ν_a^r and ν_a^f and cooling frequency ν_c (same for both, assuming $\epsilon_B^f = \epsilon_B^r$), for an electron injection spectrum $\propto \gamma^{-p}$. Top: both shocks are fast cooling. Middle: reverse shock is slow cooling, forward shock is fast cooling. Bottom: both shocks are slow cooling. The forward shock always dominates the high-frequency range, while the reverse shock, having more particles by a factor comparable to the Lorentz factor γ, produces the spectral peak.*

10.2 The Event Rate

The 'true' event rate is probably higher than the observed rate by about a factor of 500 as events appear collimated into narrow jets [144]. Frail et al. [144] estimate that the 'true' GRB event rate is about three times greater than the

estimated rate of neutron-star coalescence, and about 250 times smaller than the estimated rate of type Ibc supernova (the expected end product of a collapsar). Clearly, the collapsar scenario is capable of easily supplying a sufficient number of progenitors (including failed GRBs). Within the uncertainties of the estimates, the coalescence scenario is also (barely) capable of providing sufficient progenitors.

10.3 Offsets

The locations of GRBs in host galaxies may also put constraints on progenitor models. For instance, when a neutron star is formed in a supernova explosion it receives a substantial kick velocity of several hundred km/s (e.g., [155] and refs therein); together with the relatively long merger times ($\sim 10^8$ year) one would then expect some GRBs to be located outside the host galaxy where the binary was formed [156]. Massive star collapses, on the other hand, are expected to occur in the star-forming regions where they originated.

In a sample of Hubble Space Telescope (HST) observations it is found that GRBs consistently lie within the region of detectable rest-frame ultra-violet light of their host galaxies [157]. Since ultra-violet light is predominantly produced by young and massive stars this suggests an association with a young and massive stellar population. More quantitative: the median of the distribution of offsets of GRB sources from their host galaxy's center is less than $0.4''$ and a statistical comparison of the observed offset distribution with predicted distributions from progenitor models is consistent with a collapsar or promptly bursting binary scenario [157]; slowly bursting binary mergers such as neutron star-neutron star or neutron star - black hole systems appear inconsistent.

10.4 The Circumburst Environment

The environments of at least some bursts agree well with ordinary interstellar medium (ISM) densities [103]. These bursts thus do appear not to occur in their galaxies' halo (where one may expect a good number of slowly bursting binary mergers to reside).

Host-galaxy absorption may also provide clues to the GRB progenitor (low extinction is expected for the merger models and high extinction for massive star collapses in star-forming regions). Recently, the optical extinction toward GRBs was found to be significantly lower than one would expect from the X-ray column densities (see Fig. 22). GRBs are found typically behind large columns of gas, $N_H = 10^{22}$-10^{23} cm^{-2}, which is typical of the column densities of Galactic giant molecular clouds. This suggests that GRBs lie within star-forming regions.

The optical extinctions, however, are 10-100 times smaller than expected from the high column densities [158]. This favors theoretical findings that the early, hard radiation from GRBs and their afterglows destroys the dust in their environment, thus carving a path out of the molecular cloud through which the regular afterglow light travels relatively unobstructed [159,160]. In addition, afterglows tend to have strong MgI absorption lines, especially relative to MgII,

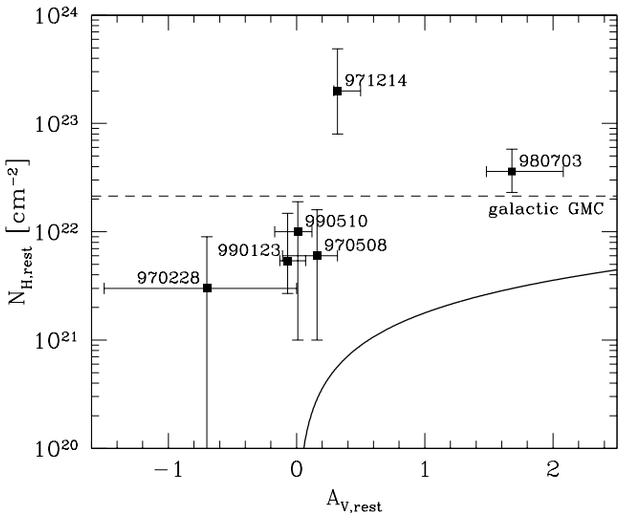

Fig. 22. *Hydrogen column density $N_{\rm H}$ vs. the optical extinction at V band A_V for a number of γ-ray burst afterglows; error bars are 68% confidence (from [158]). The solid curve is the Galactic A_V-$N_{\rm H}$ relation [161]. The dashed line shows the average column density of a giant molecular cloud of 170 $M_\odot$ pc^2 [162].*

which indicates that they originate in denser regions than the normal diffuse interstellar medium (ISM).

10.5 X-ray Lines

X-ray line features have been reported for various bursts by now, with statistical significance of $2-4\sigma$ [163,164,165,166]. However, strong upper limits were set in some other cases [167]. The relatively low statistical significance of these lines makes them somewhat uncertain, except perhaps in the afterglow of GRB 991216 [166]. The detection of these features, if true, would provide strong constraints on the progenitors and the environment, and therefore, the existence of lines is of great importance.

The lines are typically observed for a day after the event, and some of them are transient, lasting for about a day. There are two possible ways of explaining this timescale. Either there is a large amount of iron located about a light-day away from the explosion site. This iron is energized by the light from the burst and from its afterglow and then radiates either by fluorescence, recombination or thermally [168]. This mechanism is almost in contradiction with the existence of jets. If the afterglow is narrowly collimated, the iron has to be located within the opening angle of the jet (to be energized by the burst), but not directly on the line of sight in order not to block the afterglow. This setup requires fine tuning for narrow jets. The requirement of a large amount of iron came from the requirement to have a line optical depth of order unity at a relatively large

distance (a light day). Theoretical models trying to arrange such matter involve copious (but highly asymmetric) mass loss [169], a supernova explosion *prior* to the ultimate collapse of the central object [186], or reflection of very dense local ISM (whose geometry is constrained to allow the relativistic fireball to expand unimpeded [168]).

A second possibility is that iron close to the source is producing the line emission. In this case, less iron is needed due to the smaller size. In this scenario, the iron can not be energized by the burst or the afterglow, as both of these happen at large radii. Therefore, the source has to continue to operate, on a low level, for a day to energize its iron environment. The total additional energy required in this case is not large, as the overall energy observed in the line is small.

10.6 Circumstellar Wind

If at least some GRBs are produced by the core collapse of massive stars to black holes, then the circumburst environment will have been influenced by the strong wind of the massive progenitor star. For a constant wind speed the circumstellar density falls as $n \propto r^{-2}$, where r is the radial distance. In this so called circumstellar wind model, the afterglow can be described by the same synchrotron spectral shape (see Sect. 4), but with different scalings for the break frequencies ν_a, ν_m, and ν_c and the peak flux F_m (see Sect 4 and [171,172] for details).

Observationally is has turned out rather hard to detect predicted differences in the afterglow's evolution for a constant density ISM and for a circumstellar wind (see, e.g., [172,123]).

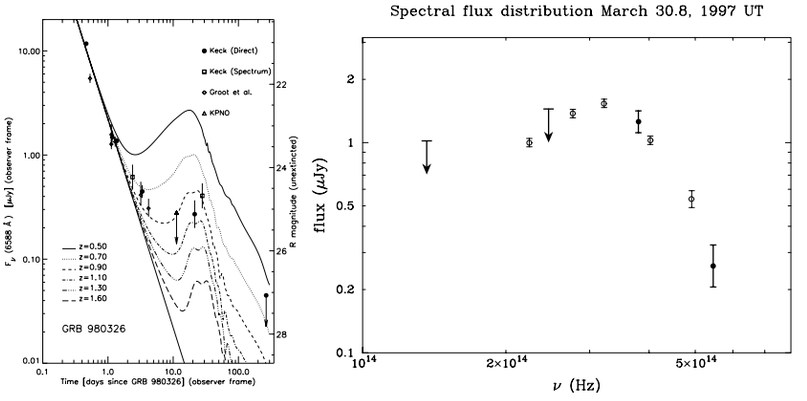

Fig. 23. *Left: R-band light curve of GRB 980326 and the sum of an initial power-law decay plus Ic supernova light curve for redshifts ranging from $z = 0.50$ to $z = 1.60$ (from [175]). Right: The broad-band spectrum of the OT of GRB 970228 at March 30.8, 1997 UT (● and upper-limit arrow). Also shown is the spectral flux distribution of SN 1998bw (○) redshifted to the redshift of GRB 970228 ($z = 0.695$). The similarity of the spectral flux distributions is remarkable (from [177]).*

10.7 The GRB/Supernova Connection

Correcting for the geometry of the explosion (i.e. taking jets into account), the total energy involved appears to cluster around 5×10^{50} ergs [144]; this is very comparable to that of supernovae (SNe).

The first evidence for a possible GRB/SN connection was provided by the discovery of SN 1998bw in the error box of GRB 980425 [173]. The temporal and spatial coincidence of SN 1998bw with GRB 980425 suggest that the two phenomena are related [173,174]. But, the clearest indication that SN 1998bw may be related to GRB 980425 comes from the fact that the radio emitting shell in SN 1998bw must be expanding at relativistic velocities, $\gamma \gtrsim 2$ [174]. This had thusfar never been observed in a SN. From minimum energy arguments, it was estimated that this relativistic shock carried 5×10^{49} erg, and could well have produced the GRB at early time [170]. Further, detailed analysis of the radio light curve [171] showed additional energy injection one month after the SN event – highly suggestive of a central engine (i.e. black hole versus neutron star formation) rather than a purely impulsive explosion.

However, GRB 980425 is most certainly not a typical GRB: the redshift of SN 1998bw is 0.0085 and the corresponding γ-ray peak luminosity of GRB 980425 and its total γ-ray energy budget are about a factor of $\sim 10^5$ smaller than those of 'normal' GRBs [173]. Such SN-GRBs may well be the most frequently occurring GRBs in the Universe. But they do not dominate the observed GRB population due to their faintness.

Evidence for a possible supernova connection for the 'normal' high-luminosity GRBs comes from the late-time red spectrum and the late-time rebrightnening of their afterglow light curves. GRB 980326 shows possible evidence that at late times the emission is dominated by an underlying supernova [175]. A template supernova light curve, provided by the well-studied type $I_{b/c}$ SN 1998bw provides an adequate description of the observations (see Figure 23).

Similarly, the optical afterglow of GRB 970228 showed indications that the standard model was not sufficient to describe the observations in detail [93]. The early-time decay of the optical emission is faster than that at later times and, as the source faded, it showed an unexpected reddening [93]. It was not until supernova-like emission accompanying GRB 980326 was found that the behavior of GRB 970228 was better understood. Also for GRB 970228 the late-time light curve and reddening of the transient can be well explained by an initial power-law decay modified at late times by SN 1998bw-like emission [176,177].

Theoretical work [178,179] has shown that these SNe most likely mark the birth events of stellar mass black holes as the final products of the evolution of very massive stars.

The relation between distant GRBs like GRB 980326 and GRB 980425/SN 1998bw is unclear. Is SN 1998bw a different phenomenon or a more local and lower energy equivalent? Attempts have been made to unify the GRB 980425/SN 1998bw phenomenon with the more distant GRBs. It has been argued that the result may be solely due to a difference in viewing angle [180,181]. And, are all afterglows consistent with such a phenomenon? Claims have been made for the

existence of SN signatures in other afterglow lightcurves, but for a number of reasons, such as the presence of contaminating host galaxy light or the brightness of the regular afterglow itself, the evidence for other cases than 970228 and 980326 is less strong. More convincing evidence may be provided by future observations of GRB afterglows around the time of the SN emission maximum, in particular a direct detection of spectroscopic features associated with SNe in a GRB afterglow: such should be possible with 8 to 10-m class telescopes such as the Very Large Telescope or Keck.

11 GRBs as Probes of the High-Redshift Universe

Key questions in modern astronomy and cosmology are: how and when did the first stars and the first galaxies form? How did structure on the largest scales in the Universe develop? How did galaxies and the star-formation rate evolve? What happened in these very early times to the intergalactic medium, to neutral hydrogen gas, and to the production of elements by stars? And, when did the epoch of re-ionization occur: the time at which the first generations of massive luminous stars ionized the hydrogen gas in the Universe?

11.1 The Relationship between GRBs and the Star-Formation Rate

Host galaxies have now been seen in most optical afterglow images (e.g, [157]). The detection of [O II] λ 3727 and Lyman α emission from some hosts indicates that these are sites of vigorous star formation. Early fits to the BATSE GRB peak flux distribution indicated redshifts $z \sim 1$ for the weakest observed bursts, but these fits relied on the assumption that the GRB rate was constant over the history of the Universe (e.g., [30]). If GRBs are related to the deaths of massive stars (whose total lifetime is very short), their rate is proportional to the star formation rate (SFR). In that case GRBs may very well be at very high redshifts, with $z \sim 6$ or greater, for the faintest bursts (e.g., [182]).

11.2 Very-High Redshifts

One of the great promises of GRB research is that GRBs can be used to probe very high redshifts ($z > 5$), because of their extreme brilliance, just as quasars are used today (up to $z \sim 5$). Lamb and Reichart [183] estimate, using current information on GRB and afterglow brightnesses from GRBs with known redshifts, that GRBs and their infrared afterglows can be detected out to redshifts of $z \sim 12$, corresponding to the first $\sim 2\%$ of the age of the Universe. Perhaps we can even detect GRBs from population III stars (the speculated very first generation of stars at $z \sim 16 - 20$ [183]). GRBs therefore promise to become a powerful tool to study the far reaches of the Universe by guiding us to regions of very early star formation, and the (proto) galaxies and (proto) clusters of which they are part.

Detection of metal-absorption lines and of the Lyα forest in a sample of GRB afterglow spectra may trace the metallicity in the Universe. And, the epoch of re-ionization can be obtained from optical and infrared spectroscopy and photometry of GRB afterglows [183]. Several models predict the re-ionization to occur at $z \sim 6 - 12$, depending on whether it was caused by quasars or the first generation of stars [184,185,186]; this is a redshift range that can be covered by GRBs. The increasingly strong evidence that GRBs are related to the deaths of massive stars (whose total lifetime is very short: only a few million years), suggests that their rate is proportional to the star-formation rate (SFR), and, by observing the rate of GRBs we can thus measure the SFR in the very high redshift Universe, out to redshifts that so far could not be reached.

Why then have very high redshift GRBs not yet been identified? This is most likely because (i) the Lyα forest and the Ly limit are redshifted to the near infrared, making the afterglows of such GRBs undetectable at optical wavelengths [183], (ii) BeppoSAX cannot localize GRBs with sufficient accuracy to make afterglow searches at near infrared wavelengths practical, and (iii) BeppoSAX triggers on brighter, so less distant, GRBs. However, HETE-2 and in particular SWIFT will be able to localize GRBs with sufficient accuracy and efficiency (detecting GRBs, respectively, ≈ 3 and ≈ 15 times fainter than BeppoSAX [187] can) that very-high redshift GRBs should be detected at greater rates than is currently possible.

12 Summary

In this review we have attempted to present the status of GRB research with a particular focus on the understanding that has resulted from the discovery of counterparts to GRBs. The answer to the 'great debate' turned out to be that GRBs occur at cosmological distances and are extremely powerful explosions. Evidence points to the existence of jets and energies in the explosion appear supernova-like. Additional evidence points to a connection with supernovae but this relation needs to be strengthened. Future observations are expected to be able to trace the afterglow light curve almost continuously from the GRB phase up to a few years after the burst. Such observations should reveal many of the ingredients discussed in this review. An example of such a light curve, is given in Figure 24. While the mystery of GRBs still needs resolution we are entering a phase in which we will increasingly start using them as tools: the extreme brightnesses allow to use them to probe the lines of sight to the GRB (providing information on the environments of GRBs and their host galaxies) and also point to the possibility to use these events to explore the high-redshift Universe.

References

1. Colgate, S.A. 1968, Canadian J. Phys. 46, S476
2. Klebesadel, R. W., Strong, I. B. & Olson, Roy A. 1973, ApJ 182, 85
3. Strong, I.B. & Klebesabel, R.W. 1976, Scientific American, October, p. 66

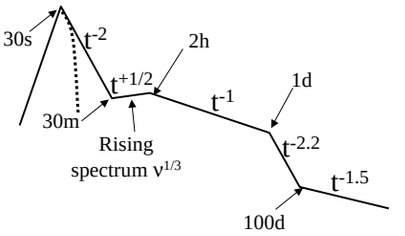

Fig. 24. *Predicted optical afterglow-light curve over more than six orders of magnitude in time (from a few seconds up to a about year). Sequentially from left to right: a rapid rise to maximum emission from the reverse shock, followed by a rapid decay $F \propto t^{-2}$ as the reverse-shock heated ejecta cool down. When the emission from the forward shock takes over, the flux is expected to rise slowly as $F \propto t^{+1/2}$ up to a second peak when the frequency ν_m of the forward shock passes the optical passband. During that rise, a positive spectral slope is expected, which has not yet been observed. The subsequent decay of roughly $F \propto t^{-1}$ is steepened to about $F \propto t^{-2.2}$ due to the jet break. The final transition is to the non-relativistic phase.*

4. Fishman, G.J. & Meegan, C.A. 1995, ARA&A 33, 415
5. Bhat, P.N. et al. 1992, Nature 359, 217
6. Schaefer, B.E. & Walker, K.C. 1999, ApJL 511, L89
7. Kouveliotou, C. et al. 1993, ApJL 413, L101
8. Paciesas, W. S. et al. 1999, ApJS, 122, 465
9. Norris, J. et al. 1995, ApJ 439, 542
10. Band, D. et al., 1993, ApJ 413, 281
11. Hurley, K. et al. 1994, Nature 372, 652
12. Sommer, M. et al. 1994, ApJL 422, L63
13. Ford, L.A. et al. 1995, ApJ 439, 307
14. Liang, E. & Kargatis, V. 1996, Nature 381, 49
15. Norris, J. et al. 1986, ApJ 301, 213
16. Link, B., Epstein, R.I. & Priedhorsky, W.C 1993, ApJL 408, L81
17. Higdon, J., Meegan, C. & Cline, T.L. 1984, AIP Conf. 141, 23
18. Hurley, K., Cline, T.L. & Epstein, R. 1984, AIP Conf. 141, 33
19. Schaefer, B.E. et al. 1998, ApJ 492, 696
20. Meegan, C.A., Fishman, G.J. & Wilson, R.B. 1985, ApJ 291, 479
21. Mazets, E.P. et al. 1981, Nature 290, 378
22. Murakami, T. et al. 1988, Nature 335, 234
23. Fenimore, E.E. et al. 1988, ApJL 335, L71
24. Palmer, D.M. et al. 1994, ApJL 35, L77
25. Band, D.L. et al. 1996, ApJ 458, 746
26. Meegan, C.A. et al. 1992, Nature 355, 143
27. Paczyński, B. 1986, ApJ 308, L43
28. Paczyński, B. 1992, Nature 355, 521
29. Paczyński, B. 1995, PASP 107, 1167
30. Mao, S. & Paczyński, B. 1992a, ApJL 388, L45
31. Brainerd, J.J. 1992, Nature 355, 522
32. Lamb, D.Q. 1995, PASP 107, 1152
33. Mao, S. & Paczyński, B. 1992b, ApJL 389, L13
34. Briggs, M.S. et al. 1996, ApJ 459, 40

35. Hartmann, D. 1995, Ap&SS 231, 361
36. Hakkila, J. et al. 1994, ApJ 422, 659
37. Hartmann, D. 1993, in *The Lives of the Neutron Stars*, Nato ASI Series, Kluwer C450, 495
38. Piran, T. 2000, Physics Reports 333, 529
39. Kommers, J.F. et al. 2000, ApJ 533, 696
40. Madau, P., Pozzetti, L. & Dickinson, M. 1998, ApJ 498, 106
41. Hughes, D.H. et al. 1998, Nature 394, 241
42. Nemiroff, R.J. 1994, Comments Astrophys. 17, 4, 189
43. Schaefer, B.E. et al. 1994, AIP Conf. Proc. 307, 382
44. Schaefer, B.E. et al. 1989, ApJ 340, 455
45. Akerlof, C. et al. 1999, Nature 398, 400
46. Goodman, J. 1986, ApJL 308, 46
47. Krolik, J. H. & Pier, E. A. 1991, ApJ 373, 277
48. Fenimore, E. E., Epstein, R. I., & Ho, C. 1993, A&AS 97, 59
49. Baring, M. G. & Harding, A. K. 1997, ApJ 491, 663
50. Lithwick, Y., & Sari, R. 2001, ApJ 555, 540
51. Mészáros, P. & Rees, M. J. 1993, ApJ 405, 278
52. Narayan, R. & Paczyński, B. & Piran, T. 1992, ApJL 395, L83
53. Rees, M. J. & Mészáros, P. 1994, ApJL 403, L93
54. Fenimore, E. E., Madras C. D. & Nayakchin, S. 1996, ApJ 473, 998
55. Sari, R., & Piran T. 1997, ApJ, 485, 270
56. Ramirez-Ruiz, E., & Fenimore, E. E. 1999, A&A, 138, 521
57. Kobayashi, S., Piran, T., & Sari, R. 1997, ApJ, 490, 92
58. Beloboradov A. 2000 ApJ 539, 25
59. Kobayashi, S. & Sari, R. 2001, ApJ 551, 934
60. Sari, R., Narayan, R. & Piran, T. 1996, ApJ 473, 204
61. Sari, R., & Piran T. 1997, MNRAS 287, 110
62. Paczyński, B. & Rhoads, J. 1993, ApJ 418, L5
63. Katz, J. I., 1994, ApJ 422, 248
64. Vietri, M. 1997, ApJ 478, L9
65. Mészáros, P., & Rees M. J. 1997, ApJ 476, 232
66. Sari, R., Piran, T. & Narayan, R. 1998, ApJL 497, L17
67. Granot, J., Piran, T. & Sari, R. 2000, ApJ 534, L163
68. Galama, T.J. et al. 1998, ApJL 501, L97
69. Waxman, E. 1997, ApJL 491, L19
70. Panaitescu A., Mészáros, P. 1998, ApJL 493, L31
71. Sari, R. 1998, ApJL 494, L17
72. Blandford, R. & Mc Kee, C.F. 1976, Phys. Fluids 19, 1130
73. Granot, J., Piran, T. & Sari, R. 2000, ApJ 513, 679
74. Gruzinov A., & Waxman E., 1999, ApJ, 511, 852
75. Granot, J., Piran, T. & Sari, R. 1999, ApJL 527, 236
76. Cohen, E., Piran T. & Sari, R. 1998, ApJ 509, 717
77. Harrison, F. A. et al. 2001, ApJ in press.
78. Sari, R. & Esin A. 2001, ApJ 548, 787
79. McNamara, B.E. et al. 1995, Ap&SS 231, 251
80. Frail, D.A. et al. 1994, ApJ 437, L43
81. Koranyi, D.M. et al. 1995, MNRAS 276, L13
82. Galama, T.J. et al. 1997, A&A 321, 229
83. Jager, R. et al. 1993, Adv. Space Res. 13, 12, 315

84. Piro, L., Scarsi, L. & Butler, R.C. 1995, Proc. SPIE 2517, 169
85. Costa, E. et al. 1997, Nature 387, 783
86. Van Paradijs, J. et al. 1997, Nature 386, 686
87. Bremer, M. et al. 1998, A&A 332, L13
88. Frail, D.A. et al. 1997, Nature 389, 261
89. Sahu, K.C. et al. 1997, Nature 387, 476
90. Metzger, M.R. et al. 1997, Nature 387, 879
91. Kulkarni, S.R. et al. 1998, Nature 393, 35
92. Kulkarni, S.R. et al. 1999, Nature 398, 389
93. Galama, T.J. et al. 1997, Nature 387, 479
94. Wijers, R.A.M.J., Rees, M.J. & Mészáros, P. 1997, MNRAS 288, L51
95. Reichart, D.E. 1997, ApJL 485, L57
96. Waxman, E. 1997, ApJL 485, L5
97. Frail, D.A., Waxman, E. & Kulkarni, S.R. 2000, ApJ 537, 191
98. Fruchter, A. et al. 2000, ApJ 545, 664
99. Galama, T.J. et al. 1998, ApJL 497, L13
100. Castro-Tirado, A. et al. 1998, Science 279, 1011
101. Piro, L. et al. 1998, A&A 331, L41
102. Wijers, R.A.M.J. & Galama, T.J 1999, ApJ 523, 177
103. Panaitescu, A. & Kumar, P. 2001, ApJ in press, astro-ph/0010257
104. Harrison, F.A. et al. 2001, ApJ in press, astro-ph/0103377
105. Panaitescu, A. 2001, ApJ in press, astro-ph/0102401
106. Berger, E. et al. 2000, ApJ 545, 56
107. Galama, T.J. et al. 2001, in preparation
108. Galama, T.J. et al. 1999, Nature 398, 394
109. Li, L. & Paczyński, B. 1998, ApJL 507, L59
110. Paczyński, B. 1993, Ann. NY Acad Sci. 688, 321
111. Mao, S. & Yi, I. 1994, ApJ 424, L131
112. Rhoads, J.E. 1997, ApJL 478, L1
113. Rhoads, J. E. 1999, ApJ, 525, 737
114. Mészáros, P. Rees, M. & Wijers, R.A.M.J. 1998, ApJ 499, 301
115. MacFadyen, A. I., Woosley, S. E. 1999, ApJ, 526, 152
116. Piran, T. 1999, in the proceedings of the Gräftåvallen workshop 'Gamma Ray Bursts: The First Three Minutes', Ed. Juri Poutanen.
117. Schmidt, M. 1999, ApJ 523, L117
118. Schmidt, M. 2001, ApJ 552, 36
119. Panaitescu, A. & Mészáros, P., 1999, ApJ, 503, 314
120. Mészáros, P., & Rees M. J. 1999, MNRAS, 299, L10
121. Moderski, R., Sikora, M., Bulik, T. 1999, astro-ph/9904310
122. Sari, R. 1997, ApJL 489, L37
123. Sari, R., Piran, T. & Halpern, J. 1999, ApJ 519, L17
124. Castro-Tirado, A. et al. 1999, Science 283, 2069
125. Fruchter, A.S. et al. 1999, ApJL 519, L13
126. Harrison, F.A. et al. 1999, ApJL 523, L121
127. Stanek, K.Z. et al. 1999, ApJL 522, L39
128. Frail, D.A. et al. 2001, ApJL in press, astro-ph/0102282
129. Wijers, R.A.M.J. et al. 1999, ApJL 523, L33
130. Rol, E. et al. 2000, ApJ 544, 707
131. Medvedev, M. V., & Loeb A., 1999, ApJ 526, 697
132. Gruzinov A. 1999, ApJ, 525, L29

133. Ghisellini, G., & Lazzati, D., 1999, MNRAS, 309, L7
134. Sari, R. 1999, ApJ, 524, L43
135. Sari, R., & Piran T. 1995, ApJ, 455, L143
136. Sari, R., & Piran, T. 1999, A&A 138, 537.
137. Sari, R. & Piran, T. 1999, ApJL 517, L109
138. Briggs, M.S. et al. 1999, ApJ 524, 82
139. Mészáros, P. & Rees, M.J. 1997, ApJ 476, 232
140. Sari, R. & Piran, T. 1999, ApJ 520, 641
141. Mészáros, P. & Rees, M. J. 1999, MNRAS 306, L39
142. Liang, E.P. et al. 1999, ApJL 519, L21
143. Kulkarni, S. R., et al. 1999, ApJ 522, L97
144. Frail, D.A. et al. 2001, in preparation
145. Akerlof, C. et al. 2000, ApJL 532, L25
146. Kehoe, R. et al. 2001, ApJ 554, L159
147. Kobayashi, S. 2000, ApJ 545, 807
148. Kumar, P., & Piran, T. 2000, ApJ 532, 286
149. Sari, R., & Mészáros, P. 2000, ApJ 535, L33
150. Eichler, D. et al. 1989, Nature 340, 126
151. Mochkovitch, R. et al. 1993, Nature 361, 236
152. Lattimer, J.M. & Schramm, D.N. 1974, ApJL 192, L145
153. Woosley, S.E. 1993, ApJ 405, 273
154. Paczyński, B. 1998, ApJL 494, L45
155. Van den Heuvel, E. & van Paradijs, J. 1998, ApJ 483, 399
156. Bloom, J.S., Sigurdsson, S. & Pols, O.R. 1999, MNRAS, 305, 763
157. Bloom, J.S., Kulkarni, S.R. & Djorgovski, S.G. 2001, AJ in press, astro-ph/0010176
158. Galama, T.J. & Wijers, R.A.M.J. 2001, ApJL 549, L209
159. Waxman, E. & Draine, B.T. 2000, ApJ 537, 796
160. Fruchter, A., Krolik, J.H. & Rhoads, J.E. 2001, ApJ in press, astro-ph/0106343
161. Predehl, P. & Schmitt, J.H.M.M. 1995, A&A 293, 889
162. Solomon, P.M., Rivolo, A.R., Barrett, J. & Yahil, A. 1987, ApJ 319, 730
163. Piro, L. et al 1999, ApJ 514, L73
164. Yoshida, A. et al. 1999, A&AS 138, 433
165. Antonelli, L. A. et al. 2000, ApJ 545, L39
166. Piro, L. et al. 2000, Science 290, 953
167. Yonetoku, D., et al, 200, PASJ 52, 509
168. Lazzati, D., Campana, S. & Ghisellini, G. 1999, MNRAS 304, L31
169. Weth, C., Mészáros, P., Kallman, T. & Rees, M. J. 2000, ApJ 534, 581
170. Tan, J. C., Matzner, C. D. & McKee, C. F. 2001, ApJ 551, 946
171. Chevalier, R.A. & Li, Z. 1999, ApJL 520, L29
172. Chevalier, R.A. & Li, Z. 2000, ApJ 536, 195
173. Galama, T.J. et al. 1998, Nature 395, 670
174. Kulkarni, S.R. et al. 1998, Nature 395, 663
175. Bloom, J.S. et al. 1999, Nature 401, 453
176. Reichart, D.E. 1999, ApJL 521, 111
177. Galama, T.J. et al. 2000, ApJ 536, 185
178. Iwamoto, K. et al. 1998, Nature 395, 672
179. Woosley, S.E., Eastman, R.G. & Schmidt, B.P. 1999, ApJ 516, 788
180. Salmonson, J. 2001, ApJL 546, L29
181. Ioka, K. & Takashi, N.T. 2001, ApJL in press

182. Wijers, R. et al. 1998, MNRAS, 294, L13
183. Lamb, D. & Reichart, D.A. 2000, ApJ 536, 1
184. Gnedin, N.Y. & Ostriker, J.P. 1997, ApJ 486, 581
185. Haiman, Z. & Loeb, A. 1998, ApJ 503, 505
186. Valageas, P. & Silk, J. 1999, A&A 347, 1
187. Weinberg N. et al. 2001, in preparation

Observations and Simulations of Relativistic Jets

José-Luis Gómez

Instituto de Astrofísica de Andalucía (CSIC), Apartado 3004, Granada 18008, Spain

Abstract. The recent improvement in VLBI arrays is providing information of the emission and magnetic field structure of relativistic jets, both extragalactic and galactic (microquasars), with unprecedented spatial and temporal resolution. These observations are revealing the importance of the hydrodynamical processes that govern the jet evolution, which can be studied by the recently developed time–dependent relativistic hydrodynamical models. Computation of the non–thermal emission from these hydrodynamical models, and its comparison with actual sources, is proving as one of the most powerful tools in the understanding of the physical processes taking place in these jets. This paper reviews some of the recent observational results, as well as the numerical models used to interpret them.

1 Introduction

Since the classical works of [19] and [13], our knowledge of the jet physics in AGNs and microquasars have improved significantly by analytical and numerical models. The analytical efforts provided the basic frame work to understand the non–thermal synchrotron and inverse Compton emission of inhomogeneous jets ([93], [72]); spectral evolution of shock waves, associated with the superluminal components ([12], [95], [63]); and polarization (e.g., [20]). The implementation of these analytical results into numerical models have allowed testing of the basic jet model hypotheses, as well as a more detailed comparison with the observations ([67], [64], [65], [43], [44], [45]). However, these early numerical models are limited by the lack of a detailed non–linear model to study the relativistic jet dynamics, being forced to adopt simplified stationary relativistic hydrodynamical models.

On the other hand, Newtonian hydrodynamical numerical models have been used to obtain a more detailed study of the jet dynamics, and its influence in the jet observational properties. These models explored with great success the morphology, dynamics and stability of jets (see e.g. reviews [109], [27]), mainly aimed to study the large scale jet structure. However, these models cannot account for the relativistic effects that are of special importance in the overall emission of jets in AGNs and microquasars.

First studies of relativistic (magneto)hydrodynamical jets were obtained for stationary flows ([136], [24], [17]). A significant step forward in the field of numerical simulations came with the development of modern high–resolution techniques in numerical hydrodynamics, mading feasible the computation of time–dependent simulations of relativistic jets ([159], [99], [100], [102], [33], [79], and reviews [101], [103], [104], [7]). These models are capable, for the first time, to

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 169–197, 2002.

study the jet dynamics with unprecedented detail, and under very similar conditions as it is thought are taking place in real sources (strong shocks, relativistic internal energies and bulk flow velocities, etc.). Some of the latest simulations have started to explore three dimensional relativistic jets ([4], [5], [6], [59]), magnetized relativistic jets ([84]), as well as jet formation and collimation making use of the first general relativity magnetohydrodynamical codes ([79], [80], [81], [105], [106]).

However, the emission structure that we observe in our VLBI images is not just a direct mapping of the jet hydrodynamical variables (pressure, density, velocity). The final radiation reaching our detectors is greatly determined by other several processes, like Faraday rotation, opacity, particle acceleration, radiative losses, and, most importantly, by relativistic effects such as light aberration and light travel time delays. For relativistic speeds (and small viewing angles) time delays can be of such importance as to render the emission images with no apparent relationship to the hydrodynamical jet structure. Therefore, the state of the art in the simulation of relativistic jets involves the computation of the emission, taking into account the appropriate relativistic and transfer of radiation processes, from the relativistic time-dependent hydrodynamical results ([46], [47], [48], [49], [66], [82], [107], [79], [5], [6], [71], [1], and review [57]). Comparison of these simulations with actual observations should provide a better understanding of the relativistic jets in AGNs and microquasars.

2 Relativistic HD and Emission Models

Most of the energy transported in relativistic jets is assumed to be carried out by a population of thermal electrons. This population determines the hydrodynamical evolution of the jet, and can be simulated by the relativistic HD codes. However, the non-thermal emission observed from these jets is originated by a second population of high energy, non-thermal particles. Detection of circular polarization in the jet of 3C 279 ([133]), as well as in 3C 84, PKS 0528+134, and 3C 273 ([60]), suggests that this non-thermal population is mainly composed by electron–positron pairs. It is still unclear how this non-thermal population is originated ([96]), perhaps by pair cascades ([92], [14]), neutron decay ([26], [42]), or by acceleration of the thermal electrons at a strong recollimation shock presumably associated with the VLBI core ([22], [98], [89], [90]). This population of non-thermal electrons is subsequently re-accelerated at shocks along the jet ([77], [38], [78], [39]), and incremented with contributions from thermal electrons accelerated at the same shocks.

In order to compute the expected emission from the hydrodynamical models is necessary to establish a relationship between the thermal and non-thermal jet populations. A common assumption considers that the particle and energy density of the non-thermal electrons is a constant fraction of the thermal electrons' ([113], [135], [46], [49], [6], [107], [79]). The population of non-thermal electrons is assumed to share the same dynamics as the thermal population, which can therefore be computed using the hydrodynamical simulations. Any exchange be-

tween internal and kinetic energy along the jet will maintain the proportionality between thermal and non-thermal populations. Only non-adiabatic processes, such as gains by particle acceleration in shocks or losses by radiation can modify this proportionality.

Radiative losses at radio wavelengths are expected to be small, except at strong shocks, such as the terminal hot spots and jet cocoon. It is therefore expected that computation of parsec scale radio emission will not be severely influenced by changes in the non-thermal population produced by radiative losses or particle accelerations. At higher energies (i.e., optical) and at sites of strong shocks it is possible to trace the non-thermal electron population gains and losses of energy by computing the electron energy transport during the jet evolution. This has been recently considered for non-relativistic magnetohydrodynamic simulations ([71]), allowing the exploration of the effects induced in the emission by synchrotron aging and electron energy gains at strong shocks.

To compute the synchrotron emission it is necessary to distribute the internal energy calculated from the hydrodynamic codes among the relativistic non-thermal electrons. This is done by assuming a power low energy distribution in the form $N(E)\mathrm{d}E = N_o E^{-\gamma}\mathrm{d}E$, with $E_{min} \leq E \leq E_{max}$, and spectral index γ. Neglecting radiative energy losses and particle accelerations, the ratio C_E between the maximum and minimum energy remains constant trough the computational domain and can be considered a free parameter of the model. The power law is then fully determined by the equation ([46])

$$N_o = \left[\frac{U\,(\gamma-2)}{1-C_E^{2-\gamma}}\right]^{\gamma-1}\left[\frac{1-C_E^{1-\gamma}}{N\,(\gamma-1)}\right]^{\gamma-2} \tag{1}$$

and

$$E_{min} = \frac{U}{N}\frac{\gamma-2}{\gamma-1}\frac{1-C_E^{1-\gamma}}{1-C_E^{2-\gamma}} \tag{2}$$

where U and N represent the electron energy density and number density, respectively, as calculated by the hydrodynamical codes.

It is still largely unknown what may be the role played by the magnetic field in the jet dynamics of AGNs and microquasars. There is some evidence pointing towards a small contribution of the magnetic field in the dynamics ([67]), although only future observations and magnetohydrodynamical simulations ([24], [161], [84], [80]) could answer this question. So far, and mainly due to the fact that the emission computations have been performed for purely hydrodynamical models, the magnetic field has been assumed to be dynamical negligible, with a magnetic energy density proportional to the particle energy density ([135]), leading to a field with magnitude proportional to $\sqrt{U}$. Once the magnetic field is considered dynamically negligible, ad-hoc magnetic field structures can be considered. To account for the small degree of linear polarization observed in many sources, the magnetic field is commonly considered to be predominantly turbulent.

2.1 Synchrotron Radiation Transfer

The transfer of synchrotron radiation have been considered in detail previously under different astrophysical scenarios, see e.g., [111], [68], [69], [64]. Its implementation for computing the polarized emission from the hydrodynamical models can be summarized as follows ([43], [44], [45], [46]).

To obtain the emission and absorption coefficients for the transfer of polarized synchrotron radiation let us consider the direction of the component of the magnetic field in the plane of the sky at a given computational cell be specified as direction 2, and let the axes 1, 2, and the direction toward the observer be directions which form a right-handed orthogonal system in that order. In this system, the emission and absorption coefficients, respectively, are then computed *in the fluid frame* using (see e.g., [111])

$$\varepsilon_\nu^{(i)} = \frac{\sqrt{3}}{16\pi}\frac{e^3}{mc^2}\, C_1^{(\gamma-1)/2}\, N_o\, (B\sin\vartheta)^{(\gamma+1)/2}\, \nu^{(1-\gamma)/2} \int_{x_{min}}^{x_{max}} x^{(\gamma-3)/2}\, [F(x) \pm G(x)]\, \mathrm{d}x \tag{3}$$

$$\kappa_\nu^{(i)} = \frac{\sqrt{3}e^3}{16\pi m}\, (\gamma+2)\, C_1^{\gamma/2}\, N_o\, (B\sin\vartheta)^{(\gamma+2)/2}\, \nu^{-(\gamma+4)/2} \int_{x_{min}}^{x_{max}} x^{(\gamma-2)/2}\, [F(x) \pm G(x)]\, \mathrm{d}x \tag{4}$$

where the plus sign is to be taken for i=1, and the minus sign is valid for i=2; ϑ is the angle between the magnetic field and the line of sight; and

$$C_1 = \frac{3e}{4\pi m^3 c^5}$$

$$x = \frac{\nu}{C_1 B \sin\vartheta E^2}$$

$$F(x) = x\int_x^\infty K_{5/3}(z)dz$$

$$G(x) = xK_{2/3}(x)$$

where $K_{5/3}$and $K_{2/3}$are the corresponding Bessel functions.

If the distribution of the magnetic field within the source is not uniform in orientation the $(1, 2)$ system will differ from cell to cell, thus it is more convenient to formulate the transfer equations in a system (a, b), which is fixed in orientation with respect to the observer. The relative orientation of the axis 2 with respect to the axis a, which defines the angle χ_B, will change from cell to cell as the magnetic field does.

The radiation field is characterized by the four Stokes parameters I, Q, U, and V, or equivalently by $I^{(a)}$, $I^{(b)}$, U, and V, where $I = I^{(a)} + I^{(b)}$ and $Q = I^{(a)} - I^{(b)}$. Provided jets in blazars exhibit very low circular polarization we can

assume $V = 0$. I is the total intensity, and Q and U determine the degree of polarization

$$\Pi = \left(Q^2 + U^2\right)^{1/2}$$

and the polarization position angle

$$\chi = \frac{1}{2}\ \arctan\left(\frac{U}{Q}\right).$$

The change of the parameters $I^{(a)}$, $I^{(b)}$ and U characterizing the radiation passing through a volume element of length ds can be obtained by solving the transfer equations in the $(1,2)$ system and transforming to the (a,b) system, given by

$$\begin{aligned}\frac{\mathrm{d}I^{(a)}}{\mathrm{d}s} = I^{(a)} &\left[-\kappa_\nu^{(1)}\sin^4\chi_B - \kappa_\nu^{(2)}\cos^4\chi_B - \frac{1}{2}\kappa_\nu\sin^2 2\chi_B\right] \\ &+U\left[\frac{1}{4}(\kappa_\nu^{(1)} - \kappa_\nu^{(2)})\sin 2\chi_B + \mathrm{d}\chi_\mathrm{F}/\mathrm{d}s\right] \\ &+\varepsilon_\nu^{(1)}\sin^2\chi_B + \varepsilon_\nu^{(2)}\cos^2\chi_B \end{aligned} \tag{5}$$

$$\begin{aligned}\frac{\mathrm{d}I^{(b)}}{\mathrm{d}s} = I^{(b)} &\left[-\kappa_\nu^{(1)}\cos^4\chi_B - \kappa_\nu^{(2)}\sin^4\chi_B - \frac{1}{2}\kappa_\nu\sin^2 2\chi_B\right] \\ &+U\left[\frac{1}{4}(\kappa_\nu^{(1)} - \kappa_\nu^{(2)})\sin 2\chi_B - \mathrm{d}\chi_\mathrm{F}/\mathrm{d}s\right] \\ &+\varepsilon_\nu^{(1)}\cos^2\chi_B + \varepsilon_\nu^{(2)}\sin^2\chi_B \end{aligned} \tag{6}$$

$$\begin{aligned}\frac{\mathrm{d}U}{\mathrm{d}s} = I^{(a)} &\left[\frac{1}{2}(\kappa_\nu^{(1)} - \kappa_\nu^{(2)})\sin 2\chi_B - 2\ \mathrm{d}\chi_\mathrm{F}/\mathrm{d}s\right] \\ &+I^{(b)}\left[\frac{1}{2}(\kappa_\nu^{(1)} - \kappa_\nu^{(2)})\sin 2\chi_B + 2\ \mathrm{d}\chi_\mathrm{F}/\mathrm{d}s\right] \\ &-\kappa_\nu U - (\varepsilon_\nu^{(1)} - \varepsilon_\nu^{(2)})\sin 2\chi_B \end{aligned} \tag{7}$$

with the average $\kappa_\nu = (\kappa_\nu^{(1)} + \kappa_\nu^{(2)})/2$. The derivative $\mathrm{d}\chi_\mathrm{F}/\mathrm{d}s$ represents the change of the plane of polarization per unit distance ds due to Faraday rotation.

A simpler formulation for the transfer of synchrotron radiation can be obtained when neglecting the different polarizations ([107]). For the total intensity, the emission and absorption coefficients can be computed using, respectively

$$\varepsilon_\nu \propto p^{(\alpha+3)/2)}\nu^{-\alpha} \tag{8}$$

$$\kappa_\nu \propto p^{(2\alpha+7)/2}\nu^{(\alpha+5/2)} \tag{9}$$

p being the thermal pressure and α the spectral index. The total intensity can then be integrated using ([19])

$$I = I_0 e^{-\tau_\nu} + \frac{\varepsilon_\nu}{\kappa_\nu}(1 - e^{-\tau_\nu}) \tag{10}$$

where τ_ν is the optical depth.

Further simplifications can be considered by ignoring opacity effects, in which case an estimation of the total intensity emission can be obtained just by adding the emission coefficient (Eq. 8) along the different cells in the line of sight ([79]).

2.2 Relativistic Effects

The presence of emitting gas at velocities close to that of the speed of light enhance the importance of the relativistic effects in the final emission structure of the simulated maps. The emission and absorption coefficients to be used in Eqs. (5-7) are those transformed into the observer's frame using the standard Lorentz transformations

$$\varepsilon^{ob}_{\nu^{ob}} = \delta^2 \varepsilon_\nu \tag{11}$$

$$\kappa^{ob}_{\nu^{ob}} = \delta^{-1} \kappa_\nu \tag{12}$$

where $\delta = \Gamma^{-1}(1 - \beta\cos\theta)^{-1} = \nu^{ob}/\nu$ is the Doppler factor; θ the viewing angle; β the flow velocity in units of the speed of light; and Γ the flow bulk Lorentz factor. Note that light aberration (see e.g., [116]) changes the orientation of the line of sight as seen in the fluid's frame, and therefore the relative orientation of the magnetic field and line of sight as seen in the fluid frame, ϑ. The emission and absorption coefficients are a function of $\sin\vartheta$ (Eqs. 3 and 4), and therefore light aberration can significantly affects the synchrotron total and polarized emission as a function of the flow velocity or viewing angle (see section 5.2).

Besides light aberration, time delay is the most important effect determining the final emission structure (no superluminal motions can be obtained from these simulations without considering the time delays between different jet regions). Provided the hydrodynamical variables are cell and time dependent, to account for delays within the jet it is necessary to compute the emission and absorption coefficients at a retarded time, given by

$$\tau = t - \frac{\overrightarrow{x}.\overrightarrow{l}}{c} \tag{13}$$

where $\overrightarrow{x}$ is the position vector of the cell, $\overrightarrow{l}$ denotes the line of sight unity vector, and c is the speed of light.

We can investigate the observational consequences of light travel delays by considering the effects produced in the emission of shocked jet material ([95], [45]). Because of the time delays between the far and near sides of a shock front, it appears rotated in the observed frame by an angle $\arccos\beta$. Depending on the pattern velocity of the shock front and viewing angle, time delays have a tendency

towards aligning the shock front to the line of sight. This may have relevant effects in the emission time variability of material being heated by a shock by producing a "phasing" effect of the emission as measured by the observed, thus allowing for very rapid variability ([117]).

Light travel delays between the forward and reverse shocks produce a lengthening of the shocked material region in the observers frame by a factor $\sin\theta/(1-\beta\cos\theta)$ ([45], [41]). Therefore, it is possible to obtain estimations of the shocked material's size in the source frame from the measured sizes, velocities, and viewing angles of superluminal components. High resolution VLBI observations ([69], [50], [52], [89], [56], [37]) reveal components sizes in some cases of the order of the jet width. If we assume commonly estimated values of $\Gamma \sim 10$ and $\theta = 10^\circ$, this implies that the emitting material associated with the superluminal component must be $\sim 1/9$ smaller than the jet width. Thus, either shocks are very thin in the source frame, or radiative losses limit the emitting region in shocks to a thin layer ([97]). It is also possible that, instead, multiple superluminal components may be associated with a single moving shock (see section 3.1).

3 Hydrodynamical Models of Superluminal Sources

Shock-in-jet models ([12], [95], [63]) have been proven to provide a general explanation for the emission variability observed in components of relativistic jets. Numerical relativistic HD and emission simulations provide a new powerful tool to improve upon these previous analytical models. With these new numerical techniques it is now possible to study with great detail the generation, structure, and evolution of strong shocks, and analyze its importance in the overall dynamical evolution and emission of jets through comparison with recent high resolution VLBI observations.

3.1 Relativistic Shocks

Superluminal components as associated with moving shock waves have been studied by relativistic hydrodynamical and emission models ([49], [79], [107]). In these models, moving shocks are induced by the introduction of perturbations in steady relativistic jets an the subsequent jet evolution is studied.

In [49] the fluid jet dynamics are computed using a relativistic, axially–symmetric jet model obtained by means of a high–resolution shock capturing scheme ([100], [102]) to solve the equations of relativistic hydrodynamics in cylindrical coordinates. The jet material is represented by an ideal gas of adiabatic exponent 4/3 and the quiescent state corresponds to a diffuse ($\rho_b/\rho_a = 10^{-3}$), relativistic ($\Gamma_b = 4$), overpressured ($p_b = 3p_a/2$), cylindrical beam with (local) Mach number $M_b = 1.69$ (subscripts a and b refer, respectively, to atmosphere and beam). The jet propagates through a pressure–decreasing atmosphere which allows the jet to expand radially. The initial pressure mismatch in the model causes recollimation shocks and expansions in the jet flow ([46]). The formation and evolution of shock waves is studied by introducing a square–wave increase

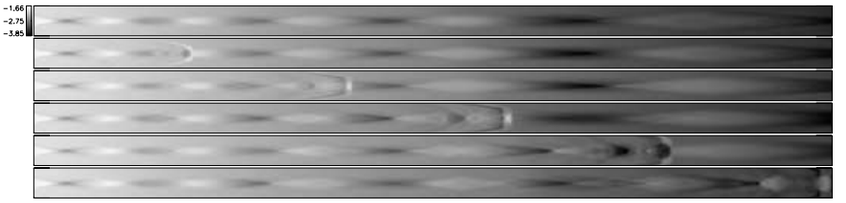

Fig. 1. Pressure distribution at six epochs (0 to 200 R_b/c in steps of 40) after the introduction of a square-wave perturbation to the flow Lorentz factor for the jet model discussed in the text. The simulation has been performed over a grid of 1600×80 cells, with a spatial resolution of 8 cells/R_b in both radial and axial directions. Reprinted from [49].

of the beam flow velocity from the quiescent value $\Gamma_b = 4$, to $\Gamma_p = 10$ during a short period of time $\tau_p = 0.75 R_b/c$. Because of the faster flow velocity in the perturbation, the fluid in front piles up, creating a shocked state, which is trailed by a rarefaction.

The resulting dynamical evolution of the perturbation along the jet is shown in Fig. 1, which contains a set of panels showing the pressure distribution at different epochs. The first panel corresponds to the quiescent jet. Both the shocked and rarefied regions in the perturbation are clearly seen. When the perturbation passes through a standing shock, the latter is "dragged" downstream for some distance before returning to its initial position as the steady jet becomes reestablished.

Figure 2 shows the total intensity maps corresponding to the stationary model (top panels), and four epochs in the evolution of the disturbance along the jet. Left and right image sequences of Fig. 2 represent the same data, but with different components identification. By looking at the unconvolved stationary total intensity image we observe a regular pattern of knots of high emission, associated with the increased specific internal energy and rest-mass density of internal oblique shocks produced by the initial overpressure in this model. VLBI cores can be interpreted as a first strong recollimation shock in the steady jet ([22], [98], [46], [89]). The regular pattern of knots should remain constant in flux and position as long as the hydrodynamical variables at the jet inlet remain unchanged. Therefore, these components resulting from the recollimation shocks may represent an alternative explanation for the stationary jet components commonly observed in many sources ([86], [134], [69], [53], [55]) as opposed to jet bendings ([2], [51]).

The time evolution of the convolved maps in Fig. 2 shows the usual core–jet VLBI structure of a blazar, with a single well–defined traveling component associated with the moving shock. The unconvolved maps show a much more complex jet structure. Due to time delays, the shocked region appears as a very extended region of higher emission (see section 2.2), which is moving and interacting with the quiescent jet. A tentative identification of components through epochs is shown in the right sequence of images of Fig. 2, where components are

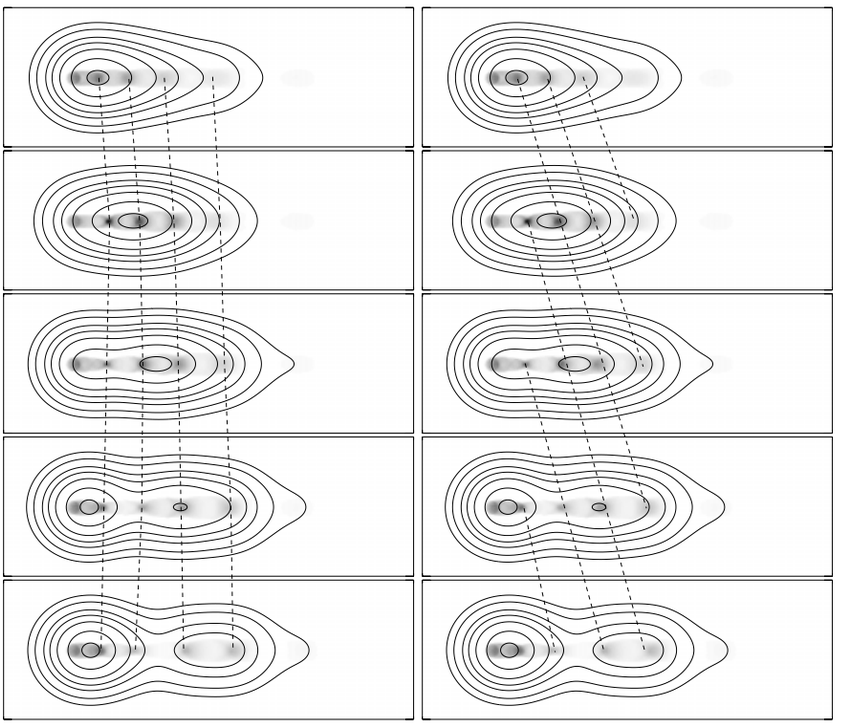

Fig. 2. Simulated total intensity maps of the hydrodynamical model presented in Fig. 1 at five different epochs. Both, left and right image sequences, represent the same data but with different components identification (see text). Grey scale (normalized to the maximum of all five epochs) shows the emission maps with the full resolution provided by the simulations. Contours show the same images once convolved with a Gaussian beam to resemble actual VLBI observations. Top panels show the stationary model. Maps are obtained for an optically thin observing frequency, and a viewing angle of 10°. Reprinted from [49].

connected by dashed lines. Without further information from the simulations, this would seem the most plausible identification of components, since it would conclude the existence of multiple superluminal components with similar apparent motions to that of the main single superluminal component obtained by analyzing the lower resolution images, that is, the convolved maps. However, this identification of components is completely wrong. When analyzing the simulations through intermediate epochs to those shown in Fig. 2 we obtain the correct identification of components, marked on the left sequence of images of Fig. 2. This shows the importance of a well time sampled monitoring when studying and identifying superluminal components through epochs. It puts in evidence how easily a wrong identification of components may result from a sparse time monitoring. Most of the information obtained from analyzing VLBI images is

deduced from the measured apparent motions, which, as shown here, may easily be completely wrong, and so the obtained conclusions.

By analyzing the structural changes in the correctly identified images of Fig. 2 we observe that the interaction of the moving shock with the underlying jet produces a temporary "dragging" of the previously stationary features, accompanied by an increase in their fluxes. Components later on come to a stop, followed by upstream motions of the inner components. This upstream motion does not represent actual upstream movement of the jet fluid, but a re-positioning of the recollimation shock closer to the jet inlet.

As the images of Fig. 2 show, detection of this predicted dragging and upstream motion of components requires high linear resolution images. Some evidence of this behavior has been found in the jet of 3C 454.3 ([98]), where 43 GHz VLBA observations have revealed the existence of a stationary component that moves downstream slightly before returning back upstream as a moving component passes it. Other evidence has been found in the jets of 3C 120 ([50], [54]), 0735+178 ([30]), 3C 279 ([134]), and may be expected in other sources as more high–frequency images become available.

In [107] the appearance of VLBI knots is studied by obtaining the total intensity emission from relativistic flows computed using the relativistic hydrodynamical code of [33]. Computation of the synchrotron radiation is obtained by computing the emission and absorption coefficients (Eqs. 8 and 9), taking into account opacity effects to integrate the transfer equation (Eq. 10). Time delay effects are ignored because the jet structures are found to move at barely relativistic speeds.

Making use of this numerical model, perturbations in the jet are studied in [107] by introducing a sinusoidal modulation of the inflow Lorentz factor between 1 and 10. Figure 3 shows the obtained density plots before, and after the perturbations are introduced. The relative dominance of the intrinsic emissivity and Doppler boosting in the intensity images is studied by computing the emission at different observing viewing angles. For small viewing angles the image morphology is found to be determined primarily by the Doppler boosting of the high-velocity jet, whereas at larger angles the intrinsic emissivity is more important. Blazars are assumed to be observed along small viewing angles, and therefore the appearance of VLBI knots is determined primarily by the Doppler boosting of fast moving jet perturbations.

3.2 Trailing Shocks

The evolution of a strong shock wave cannot ideally be isolated from the underlying jet flow. During its motion along the jet the shock wave interacts with the ambient jet medium, as well as the quiescent flow. This highly non–linear interactions trigger a local pinch instability ([59]) that leads to the formation of a series of conical shocks. Some of these shocks are present in the simulations of Fig. 1 and have been studied in detail by [1].

Figure 4 shows the Lorentz factor distribution for a jet simulation after the passage of a strong shock. Multiple conical recollimation shocks ("trailing

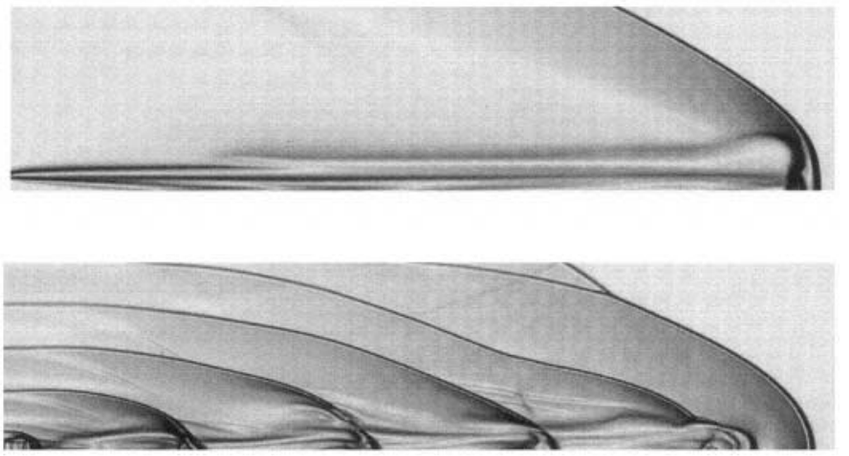

Fig. 3. Schlieren-type images of laboratory frame density gradient for a jet with a Lorentz factor of 10 and adiabatic index of 4/3. Bottom image shows the same jet after the inflow Lorentz factor has been sinusoidally modulated between 1 and 10 to induce perturbations. Reprinted from [107]

.

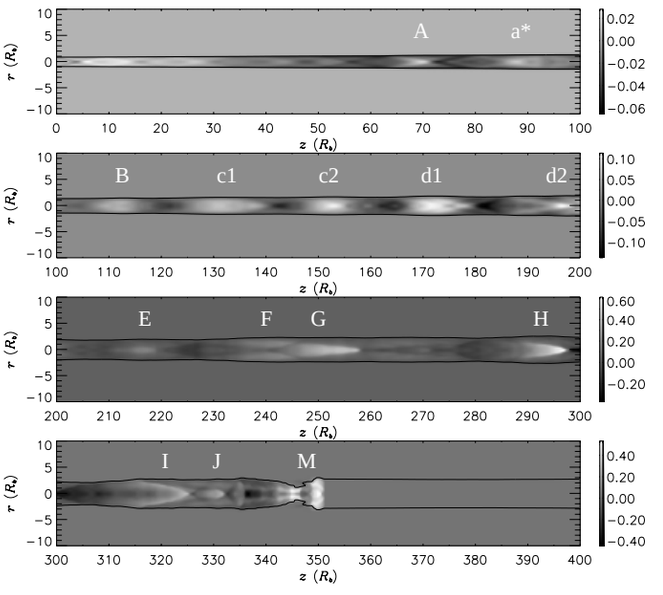

Fig. 4. Relative variation with respect to the quiescent jet of the Lorentz factor (logarithmic scale). Multiple conical recollimation shocks ("trailing shocks") are found to follow the main shock labeled "M". Reprinted from [1].

shocks") can be found to follow the main perturbation. Although their strength is a function of the distance from the jet inlet, simulated total intensity maps show that they should be strong enough as to be detectable by present VLBI arrays ([1]).

These trailing shocks can be easily distinguished because they appear in the simulated maps as components being released on the wake of a primary superluminal component (associated with the leading shock), instead of being ejected from the core of the jet. Those trailing components appearing closer to the core show small apparent motions and a very slow secular decrease in brightness, from which they could be identified as stationary components. Those appearing farther downstream are weaker and can reach apparently superluminal motions. Their oblique nature should also result in polarization properties different from that of the main planar leading shock. The existence of these trailing components indicates that not all observed components necessarily represent major perturbations at the jet inlet; rather, multiple emission components can be generated by a single disturbance in the jet.

A sample of 42 γ–ray blazars observed at high frequencies with the VLBA has revealed that stationary components are more common than previously thought ([75]). In 27 of those sources at least one non–core stationary component has been observed. By analyzing the properties of these stationary features two different classes of stationary components are established ([75]): those within about 2 mas[1] of the core, probably associated with standing hydrodynamical compressions, and those farther down the jet, probably associated with bends in the jet. These inner stationary features are in good agreement with the properties predicted for the trailing shocks, and therefore their association seems a plausible interpretation for their nature. Polarimetric high resolution VLBI observations should provide the necessary information as to confirm or rule out this hypothesis.

3.3 Jet Instabilities and the Formation of Knots

Relativistic jets in AGNs and microquasars are thought to be subject to instabilities, perhaps due to changes in their feeding from unstable accretion disks. These jet instabilities have been studied with great detail by linear stability analysis of the linearized fluid equations and by non–linear hydrodynamical simulations (e.g., [16], [102], [58], [115], [59], [137]).

Numerical simulations by [137] have revealed that mode–mode interactions in 3D, such as helical surface and helical body mode interactions and coupling to pinch modes, may lead to the formation of relatively stationary knots along the jet beam. In particular, wave–wave interactions are shown to lead to the formation of knots internal to the jet beam which are nearly stationary close to the jet inlet, but move and develop shock spurs at larger distances. These mode–mode interactions, as well as the trailing shocks, may explain some of the

[1] $1ma = 10^{-3} arcsec \equiv 4.85 \times 10^{-9} rad \equiv 2.77 \times 10^{-7} deg$ - *the eds.*

puzzling knots evolution observed in the galactic superluminal GRO J1655-40 ([61]).

4 Magnetic Fields in Relativistic Jets

Although recent polarimetric VLBI observations are providing added information on the magnetic field strength and structure at different jet scales, the role played by the magnetic field in the jet dynamics is still largely unknown. In order to have a dynamically important magnetic field we should look for jet regions where the magnetic pressure $B^2/8\pi$ dominates over the thermal jet plasma pressure. This could be found in the inner jet regions, where magnetic pressure should be of importance for the initial jet formation and collimation.

4.1 Formation, Collimation, and Acceleration of Jets

Observation of the inner jet regions, where jets are formed, collimated and accelerated, requires the highest possible linear resolution in terms of black hole Schwarzschild radii, which determines the scale length for the system. It is therefore in nearby sources with known massive central black hole where high frequency VLBI observations can provide the necessary linear resolution. This has been achieved by global 43 GHz VLBI observations of the jet in M87 ([74]), revealing that strong collimation of the jet takes place at 30-100 Schwarzschild radii (r_s) from the black hole, continuing out to $\sim 1000\ r_s$.

Thanks to the development of recent general relativistic magnetohydrodynamic (GRMHD) numerical codes ([79], [80], [81]) it is now possible to study the production of relativistic jets by numerical simulations (see e.g., reviews by [105], [106]). The common scenario for jet production requires a differentially rotating accretion disk surrounding the massive central object. The disk is also threaded with an axial magnetic field of sufficient strength to exert a braking force on the rotating plasma, removing angular momentum and transfering it along the magnetic field lines. These rotating magnetic twists push out and pinch the plasma into a jet. This sweeping pinch mechanism appears to be nearly universal ([106]).

Numerical GRMHD simulations of jet formation in a rapidly rotation Kerr black hole have been performed for the cases of a co–rotating and counter–rotating Keplerian accretion disk ([81], Fig. 5). For the co–rotating disk case, a pressure driven jet is formed by a shock in the disk, together with a weaker magnetically driven jet outside the pressure driven jet. However, for the counter–rotating disk case, a powerful magnetically driven jet is formed inside the pressure driven jet. This magnetically driven jet is accelerated by a strong magnetic field created by frame dragging in the ergosphere of the black hole.

4.2 Intrinsic Polarimetric Differences in Jets of AGNs

Polarimetric VLBI observations have revealed intrinsic differences in the jets of BL Lacertae type objects and QSOs that cannot be explained solely by dif-

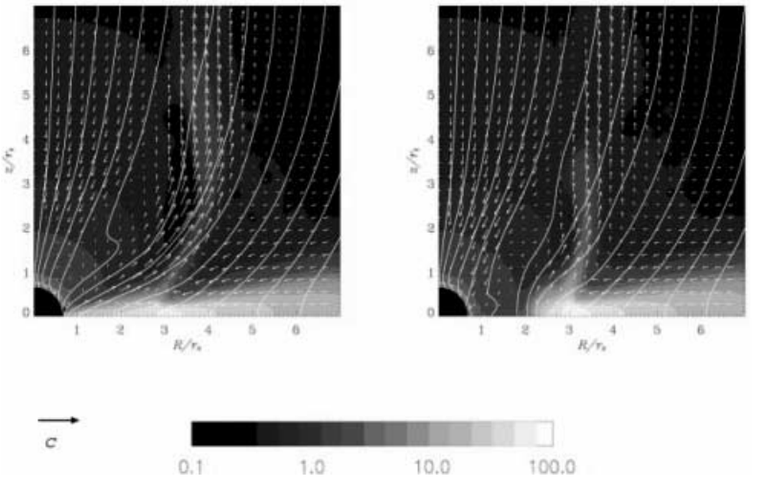

Fig. 5. Numerical models of jet formation for the case of a counter–rotating (left) and co–rotating disk (right). Grey scale shows the logarithm of the proper mass density; vectors indicate velocity; solid lines show the poloidal magnetic field. Reprinted from [81].

ferences in the viewing angle to the flow axis. First evidence for these intrinsic differences were observed in the polarization properties of the jets in the Pearson–Readhead sample through VLBI observations at 6 cm ([29], [21]). These observations have shown that the magnetic fields in BL Lac jet components are commonly perpendicular to the jet structural axis, while for QSOs the orientation is typically aligned to the jet axes. Recent observations ([31], [35], [34]) confirm these differences, but also provide evidence that about 30% of the BL Lac sources in the sample present aligned magnetic fields, similar to those found in QSOs. This difference in the polarization properties of BL Lacs and QSOs is interpreted by associating the observed knots with moving transverse shocks in jets containing mainly tangled magnetic fields. Shocks would be stronger and more commonly observed in BL Lacs, leading to the observed perpendicular fields in the knots. On the contrary, QSOs would be required to be less active, with weaker shocks that would never dominate in polarization.

This larger activity in BL Lacs is also supported by the University of Michigan long–term total and polarization monitoring program ([3]). BL Lacs are found to be more highly variable in total flux than QSOs, and to present quasi simultaneous variations at different frequencies. This also suggests the existence of intrinsic opacity differences between the two classes of objects. The analysis of the polarized light curves is indicative of the existence of propagating shocks during outbursts. The larger variability in BL Lacs then supports the model in which shock formation is more frequent in BL Lacs parsec-scale jets than in QSOs.

The different activity in these two classes of objects can be interpreted in terms of jet instabilities ([3]). In [115] it is found that higher jet stability can be obtained in faster and colder jets. However, simulations by [102] show that highly supersonic jets (those in which the kinematic relativistic effects due to high Lorentz factors dominate) present a rich internal structure, with multiple internal shocks and extended overpressured cocoons. Both set of simulations ([115], [102]) agree on finding hot jets (i.e., beams with internal energies comparable to the rest–mass energies) the most stable. Further relativistic HD simulations of jet stability are required to explore the parameter space to determine in which cases jets are expected to be more or less stable, and then establish a relationship with the jets in QSOs and BL Lacs (presumably less stable).

It is also possible that these apparent differences in the jet stability of BL Lacs and QSOs depend on the jet scales studied. Higher resolution (1.3 cm and 7 mm) polarimetric VLBI observations ([89], [90]), therefore exploring inner jet regions than those mapped at 6 cm, reveal no significant differences in the polarization properties of BL Lacs and QSOs. Furthermore, comparison between radio and optical reveals a strong correlation in the polarization of the radio core and overall optical polarization of the source, suggesting a common and possibly co–spatial origin for the emission at theses frequencies. Magnetic fields perpendicular to the flow direction are commonly observed for the radio cores. Similar orientations are found in the optical, suggesting that the emission at both wavelengths is originated by a strong transverse shock, perhaps the first recollimation shock in the jet, associated with the radio core (see section 3.1).

Although no significant differences in the polarization of BL Lac and QSOs are found in the high resolution observations of [90], the previous dichotomy is translated to high– and low–optically polarized compact radio–loud quasars (HPQs and LPRQs, respectively). LPRQs are found to have components with magnetic fields predominantly parallel to the jet, while in HPQs components tend to have perpendicular magnetic field orientations. This is interpreted assuming that LPRQs represent a quiescent phase of blazar activity, in which the inner jet does not contain strong moving shock waves.

4.3 Intraday Polarization Variability

Rapid variations, with time scales less than a day, in both total and polarized flux density have been observed in several radio sources (see e.g. review by [130]). If intrinsic and resulting from incoherent synchrotron radiation, this intraday variability (IDV) implies jets with bulk Lorentz factors between approximately 30 to 100, larger than the largest values inferred from superluminal motions, and requiring implausibly high brightness temperatures ([15]). Although most IDV at radio wavelengths probably includes some contribution from propagation effects ([114]), recent polarimetric VLBI observations reveal that some of these variations may be intrinsic to the sources ([32], [33], [36]).

One of the first sources found to exhibit IDV is the BL Lac object 0716+714. VLA observations of this source revealed a rotation of the polarization angle by about 50° in 12 hours. By comparison with simulations VLBI observations, it

was possible to determine that, contrary to what it was expected, the region responsible for this variability was not the core, but probably a feature located at about 25 milliarcseconds from it ([32]). Further IDV in polarization, but *not* in total flux, have been found in the inner regions of the jets in several other sources, including 0917+624, 0954+658, 1334-127, 2131-021, and 2155-152 ([33], [36]). In the case of 2155-152, IDV variations were seen directly in the polarized intensity images of this source at 5 GHz, where only one of the two polarized milliarcsecond scale features varied. This represents one of the first evidence that IDV in polarization is intrinsic to the source. Although propagation of shocks through turbulent jets may explain some of the observed IDV properties ([95]), further observations and theoretical modeling are necessary to obtain a more detailed picture of the jet physical processes required to explain the exhibited IDV.

5 Jet Environments

Propagation of jets is greatly determined by the distribution of gas in their host galaxies. As the same time, the interaction of the jet with the ambient gas may play an important role in determining some of the observational properties of the emission-line gas. Distinct signs of interaction between a collimated radio jet and a clumpy Narrow Line Region (NLR) are commonly found in the form of morphological associations between radio and optical structures ([18], [28], [10], [129]). The radio-optical association suggests that the interaction of the jets with the interstellar medium strongly influences the dynamics of the ionized gas in the NLR. Furthermore, the ambient gas can be ionized by the direct interaction with the jet bow shock, or by diffuse photoionizing radiation fields produced in the shocks generated by such interactions, as observed in 3C277.3 and 3C171 ([122]).

Exploration of the time-dependent interaction of jets with the NLR have been performed by numerical non-adiabatic hydrodynamical simulations ([119]). These simulations show that the association between the radio and optical emission can be explained as a natural consequence of the expansion of a hot jet cocoon into the interstellar medium. Radiative losses create an envelope of dense cool gas and discrete emission-line knots which can be associated with the narrow-line clouds themselves. Some of these clouds might be partially neutral and represent sites of jet-induced star formation ([119]). Simulated H_α emission shows similar total line widths to those observed in NLR of Seyfert galaxies, presenting large-scale variations in the radial velocities of the clouds due to the stratified pressure in the bow shock region of the jet ([120]).

Direct collisions between the jet and clouds of the BLR and NLR are statistically expected, depending on the assumed values for the cloud sizes and the filling factor (e.g., [92]). Three-dimensional numerical hydrodynamical simulations ([132], [128]) have been used to study direct collisions of a jet with isolated clouds. These simulations show that, although powerful jets would disperse the clouds, for off-center collisions nonaxisymmetric instabilities are induced in the

jet and can eventually disrupt it. These interactions could explain some of the morphologies observed in compact steep-spectrum sources, such as the strongly bent geometries found in some of these sources ([92]).

5.1 Jet-Cloud Collisions in 3C 120

The radio galaxy 3C 120 was one of the first four sources in which superluminal motion was detected on the scale of parsecs ([118]) to tens of parsecs ([131]). High resolution polarimetric VLBI observations ([50], [52]) have revealed a richer, more rapidly changing structure in total and linearly polarized intensity than that found in other relatively nearby compact extragalactic jets ([74], [125], [76]). Thanks to its proximity (z=0.033), millimeter VLBI observations allow to probe the inner jet structure of 3C 120 with very high linear resolution, $\sim 0.1\ h_{65}^{-1}$ pc at 43 GHz. This provides enough resolution as to test some of the predictions obtained with the numerical simulations outlined in section 3. Towards this aim, the radio galaxy 3C 120 has been studied with unprecedented spatial and temporal resolution by performing a 16 epoch monthly monitoring with the VLBA at 22 and 43 GHz in dual polarization. This represents the most thorough study of a relativistic jet to date, complete with the highest resolution and polarization ([56]). The obtained sequence of images at 22 GHz is reproduced from [56] in Fig. 6.

The images show the appearance of a new strong superluminal component, labeled "O" in Fig. 6, coincident with a major outburst in the light curve. The passage of this new superluminal component triggered the appearance of a stationary feature ("M" in Fig. 6) that presents also enhanced linearly polarized emission. This behavior is in agreement with the numerical simulations of "trailing shocks" by [1], which explain the appearance of this stationary feature as a consequence of the jet instabilities produced by the passage of the strong leading shock, which would be associated with component "O".

Figure 7 shows the light curves for several of the components found in 3C 120 where a remarkable brightening can be observed starting when components reach a distance from the core of ~ 2 mas. The most pronounced (in terms of change in flux density) flare corresponds to the component labeled "L", which increased its total flux density by a factor of 9, becoming the strongest feature in polarized intensity (Fig. 6). This flare is accompanied by rotation of the magnetic vector and an increase in degree of polarization at both 22 and 43 GHz. The slower rotation of the magnetic vector at 43 GHz reveals a progressive increase in the rotation measure (RM) of component "L", reaching a value of $\sim 6000 \pm 2400$ rad m^{-2} at peak emission.

The rapid flares in the flux densities of components are followed by equally fast declines when they reach ~ 3 mas from the core. It is difficult to explain such rapid changes in the total and polarized flux density, as well as polarization angle, for components located between a deprojected distance of 4 and 10 pc from the core. The bending of the jet appears to be too slight to cause such variations in brightness (from changing relativistic beaming of the radiation relative to the observer) without an accompanying acceleration of the apparent

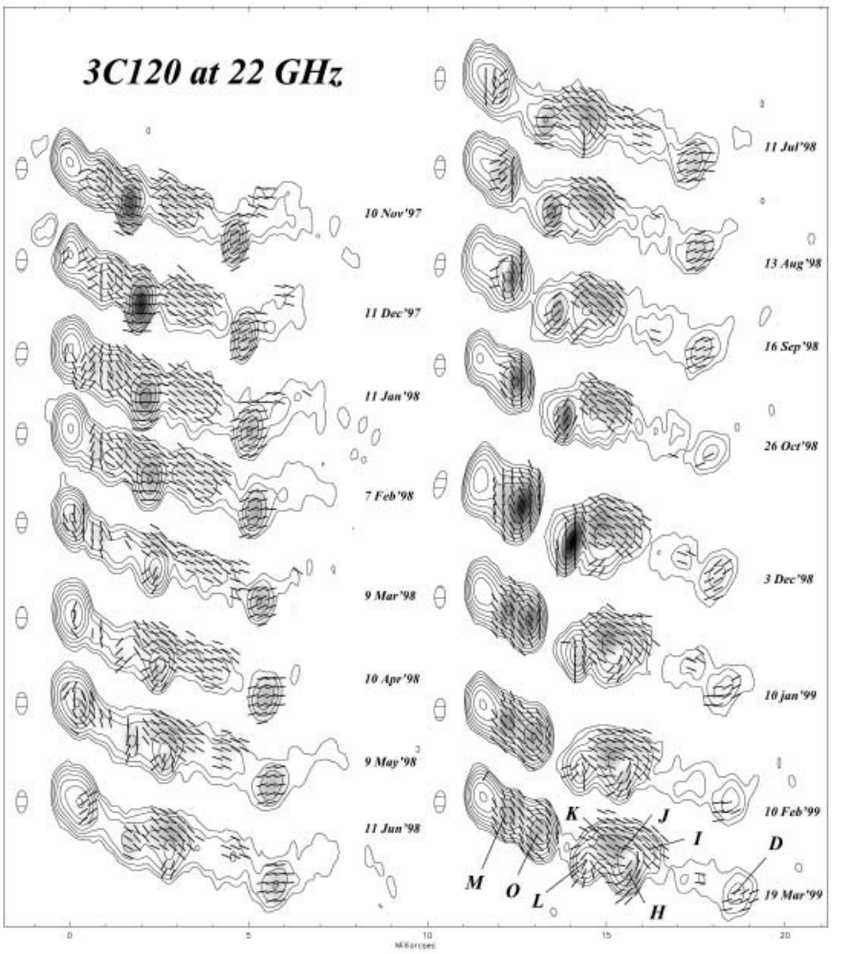

Fig. 6. 16 epoch monthly monitoring of the jet in the radio galaxy 3C 120 obtained with the VLBA at 22 GHz. Total intensity is plotted in contours, linearly polarized intensity in gray scale, and magnetic vectors with bars. Reprinted from [56].

velocity and presence of a stationary component at this site ([2], [45]). Rather, interaction between the jet and a dense cloud in the external medium seems the most plausible explanation. Similar interactions between the jet and interstellar medium were inferred previously from the discovery of an emission-line counterpart to the more extended radio jet in 3C 120 ([62], [9]). It appears that this interaction is most intense along the southern border of the jet, where the gentle northward curvature causes components with higher than average momentum to collide with the external medium or cloud. Indeed, it is at the beginning of this bend that component "L", which is the closest to the southern jet border and

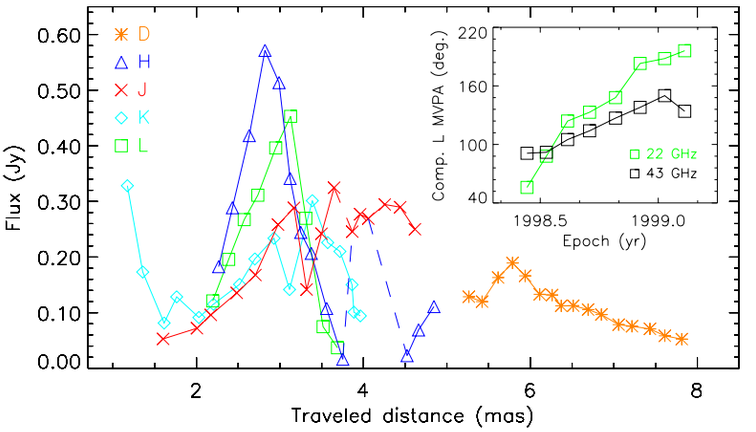

Fig. 7. Light curves for components in the radio galaxy 3C 120 shown in Fig. 6. Inset panel shows the evolution with time of the magnetic vector position angle of component "L". Reprinted from [56].

the one exhibiting the largest flare, began to increase its flux density. This behavior is explained if the magnetic field and population of relativistic electrons in component "L" were enhanced by the shock wave produced by interaction of the jet with the external medium, resulting in a rapid rise in synchrotron emission. The observed increase in the degree of polarization is then explained as a consequence of ordering of the field by the shock wave. The rotation of the magnetic vector observed in component "L" can be interpreted as Faraday rotation, the level of which can be estimated from the different polarization angles observed at 22 and 43 GHz (Fig. 7). After removing this effect, the relative orientation of the magnetic field and velocity vector (which rotates as the component follows the bend in the jet) remains at $40\pm10^{\circ}$. The observed Faraday rotation can be explained by an ionized cloud along the line of sight that may also physically interact with the jet.

For a cloud at a temperature of 10^4K, free-free absorption provides an estimated electron density of $\sim 5 \times 10^4$ cm^{-3}. The observed RM of $\sim 6000 \pm 2400$ rad m^{-2} would then require a magnetic field strength of ~ 0.4 mG. Similarly large RMs have been found in several extragalactic jets ([126], [124], [73]), with estimated magnetic fields of the same order. This electron density and distance from the central engine is intermediate between those of the broad and narrow emission–line clouds in AGNs. Given its high column density, $\sim 6 \times 10^{22}$ cm^{-2}, such a cloud could be detected in absorption if there is a substantial neutral atomic or molecular component, which is expected. Such an observation, which could be carried out with the VLBA in spectral-line mode, would determine the radial velocity of the cloud and therefore whether it is moving toward or away from the central engine.

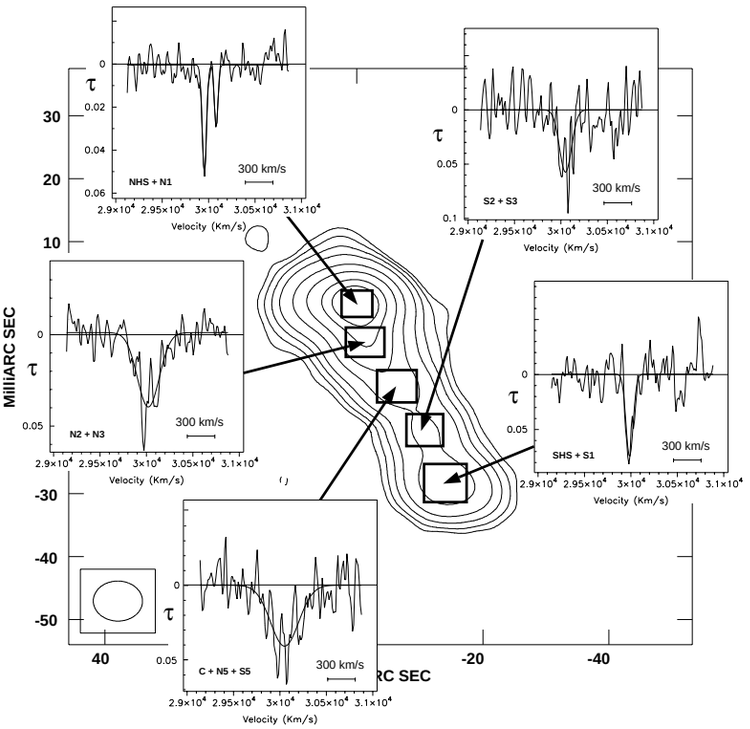

Fig. 8. Integrated profiles of HI absorption in five regions of the compact symmetric object 1946+708. Contours show the radio continuum image at 1.29 GHz. Reprinted from [112].

Spectral–line VLBI observations have allowed to investigate the ambient medium in AGNs with great detail ([108], [123], [40], [112], [110]). The common scenario outlined by these observations consists of an accretion disk or torus surrounding the central engine of the AGN. Depending on the source geometry, part of the jet radio continuum would be absorbed by the atomic gas that mainly comprises the disk, producing the HI absorption lines. The UV photons from the central engine would ionize the inner gas of the AGNs, leading to free-free absorption of the jet radio continuum.

In the compact symmetric object 1946+708, VLBA spectral–line observations ([112]) have revealed narrow HI absorption lines in the jet northern hot spots, indicating small clouds of warm neutral gas associated with an extended clumpy torus located between the radio jet and the observer (see Fig. 8). The high velocity dispersion and column density toward the core of 1946+708 suggests fast moving material, possibly in rotation around the central engine.

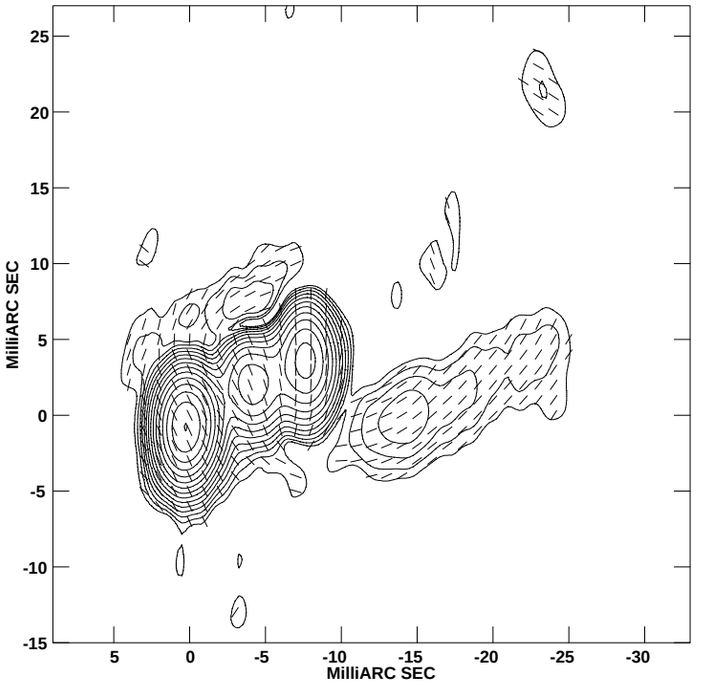

Fig. 9. Linear polarization distribution of the blazar 1055+018 obtained with the VLBA+Y1 at 5 GHz. The ticks show the magnetic vector orientation. Reprinted from [8].

5.2 Jet Stratification

Propagation of relativistic jets through an ambient medium leads to the formation of shear layers. Such layers have been invoked in the past to account for a number of observational characteristics observed in FR I ([85], [88]) and FR II sources ([121]).

One of the best observational evidence for these shear layers have been recently found in the arcsec scale jet of 1055+018 ([8]). Polarization imaging with the VLBA at 6 cm shows that 1055+018 apparently consists of i) an emission spine along the jet axis containing a series of knots in which the magnetic field is predominantly perpendicular to the axis, and ii) a boundary layer in which the magnetic field is predominantly parallel to the axis, as shown in Fig. 9. The aligned magnetic field in the shear layer is assumed to be originated by the jet interaction with the ambient gas. However, this cross-section asymmetry presents a rather unusual structure, since it is observed to change with distance along the jet: the shear layer is only visible on one side of the jet at a time.

Three dimensional numerical hydrodynamic and emission simulations have been performed to investigate the formation of shear layers and their implications

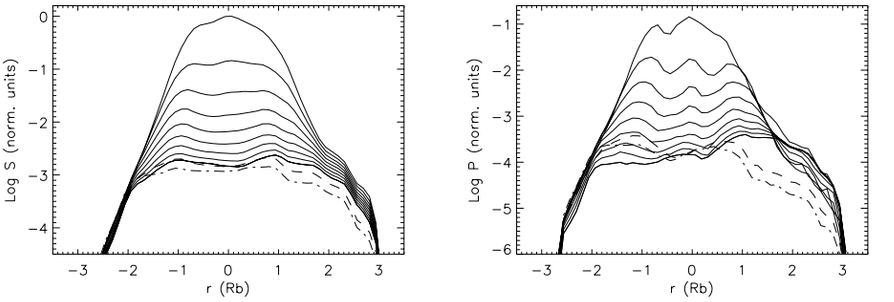

Fig. 10. Logarithm of the integrated total (left) and polarized (right) intensity across the jet for different viewing angles. Lines are plotted in intervals of 10° between an angle of 10° (top line in both plots) and 90° (showing a progressive decrease in emission). Dashed (dot dashed) lines correspond to an observing angle of -130° (-170°). Reprinted from [6].

for the jet emission ([6]). These simulations show that the interaction of the jet with the external medium gives rise to a jet stratification in which a fast spine is surrounded by a slow high-energy shear layer. In order to explore the polarization observational properties of such a jet stratification, [6] considered an ad hoc distribution of the magnetic field consisting of two components: i) a toroidal field present both in the jet spine and the shear layer, and ii) a second component aligned in the shear layer and radial in the jet spine. The resulting projected magnetic field is aligned in the shear layer and is perpendicular in the jet spine, as suggested by several observations ([85], [121], [8]).

Because of this helical magnetic field structure in the shear layer, an asymmetry in the emission is found to appear across the jet. This asymmetry is more pronounced in the polarized emission, and is a function of the viewing angle, as shown in Fig. 10. The synchrotron radiation coefficients are a function of the sine of the angle between the magnetic field and the line of sight in the fluid frame, ϑ (Eqs. 3 and 4). Therefore, asymmetries in the distribution of ϑ will be translated into the emission maps, giving rise to the jet asymmetry. In order to compute ϑ is necessary first to Lorentz transform the line of sight from the observer's, θ, to the fluid's frame θ' (see e.g., [116]). For a helical magnetic field with a pitch angle ϕ, measured with respect to the jet axis, the angles ϑ^t and ϑ^b (where superscripts t and b refer to the top and bottom of the jet, respectively) increase by 2ϕ (note that $\vartheta^{t,b}$ is always defined as positive). Therefore, as long as ϕ is different from zero or $\pi/2$, i.e. the field is neither purely aligned nor toroidal, the factor $\sin\vartheta^{t,b}$ in the synchrotron radiation coefficients will introduce an asymmetry in the jet emission. This asymmetry will reach a maximum value for a helical magnetic field with $\phi = \pi/4$. However, indistinctly of the helix pitch angle, the predominance of $\sin\vartheta^t$ or $\sin\vartheta^b$ will reverse at $\theta' = \pi/2$, respectively, which corresponds to a viewing angle in the observer's frame of $\cos\theta_r = \beta$. For a helical field oriented clockwise as seen in the direction of flow motion (i.e., the aligned component of the field is parallel to the jet flow), and

for $\theta' < \pi/2$ the bottom of the jet will show larger emission, while for $\theta' > \pi/2$ the top of the jet will be brighter (the opposite is true for a helical field oriented counter-clockwise, i.e. $\phi > \pi/2$). The maximum asymmetry will be obtained for $\theta' = \phi$ and $\theta' = \pi - \phi$, and the fastest transition (with changing θ') between top/bottom emission predominance will be obtained for ϕ close to $\pi/2$, i.e. when little aligned field is present.

It is interesting to note that for $\theta \sim \theta_r$, small changes in the jet velocity or the viewing angle will produce a flip in the top/bottom jet total and polarized emission dominance. This model has been used by [6] to interpret the shear layer structure observed in 1055+018 ([8]). For this, 1055+018 is required to be oriented close to θ_r and to contain a shear layer with a helical field. In this case, the flip in the top/bottom orientation of the polarization asymmetry in 1055+018 is interpreted as due to a jet deceleration, as observed for several components in this source ([8]).

6 Conclusions

Numerical relativistic (magneto)hydrodynamic and emission simulations have proven to be a powerful tool to understand the physics of jets in AGNs and microquasars through direct comparison with observations. These models are capable of study the jet dynamics with unprecedented detail, and under similar conditions to those in actual sources (relativistic internal energies and bulk flow velocities). Computation of the non–thermal radio emission allows to study the relationship between radio knots and internal shock waves. These simulations show that the evolution of moving shock waves is greatly determined by their interaction with other, standing, shock waves, as well as the underlying jet flow and external medium. "Upstream" knot motions, "dragging" of previously stationary components, and formation of multiple "trailing" components after the passage of a major strong shock are some of the predictions obtained from these simulations. First observational evidence of these features are being obtained thanks to the recent millimeter polarimetric VLBI observations.

The dynamical and emission evolution of jet components may be severely affected by interactions with the external medium. An extensive monitoring of the radio galaxy 3C 120 with the highest resolution and in polarization has provided direct imaging of the interaction between jet components and the external medium, resulting in rapid changes in the total and linearly polarized emission of components. These interactions between jet and ambient medium may also result in a jet stratification, in which a fast spine is surrounded by a slow high–energy shear layer, leading to an emission cross section jet asymmetry.

Further numerical simulations, and its comparison with high resolution observations should provide new insights towards the understanding of the physical processes taking place in the jets of AGNs and microquasars.

References

1. Agudo, I., Gómez, J. L., Martí, J. M., Ibáñez, J. M., Marscher, A. P., Alberdi, A., Aloy, J. M., & Hardee, P. E. 2001, ApJ, in press
2. Alberdi, A., Marcaide, J. M., Marscher, A. P., Zhang, Y. F., Elosegui, P., Gómez, J. L., & Shaffer, D. B. 1993, ApJ, 402, 160
3. Aller, M. F., Aller, H. D., Hughes, P. A., & Latimer, G. E. 1999, ApJ, 512, 601
4. Aloy, M. A., Ibáñez, J. M., Martí, J. M., & Müller, E. 1999, ApJS, 122, 151
5. Aloy, M. A., Ibáñez, J. M., Martí, J. M., Gómez, J. L., & Müller, E. 1999, ApJ, 523, L125
6. Aloy, M. A., Gómez, J. L., Ibáñez, J. M., Martí, J. M., & Müller, E. 2000, ApJ, 528, L85
7. Aloy, M. A., & Martí, J. M. 2001, this proceedings
8. Attridge, J. M., Roberts, D. H., & Wardle, J. F. C. 1999, ApJ, 518, L87
9. Axon, D. J., Unger, S. W., Pedlar, A., Meurs, E. J. A., Whitte, D. M., & Ward, M. J. 1989, Nature, 341, 631
10. Best, P. N., Carilli, C. L., Garrington, S. T., Longair, M. S., & Röttgering, H. J. A. 1998, MNRAS, 299, 357
11. Blandford, R. D., & Rees, M. J. 1974, MNRAS, 169, 395
12. Blandford, R. D., & McKee, C. F. 1976, PhFl, 19, 1130
13. Blandford, R. D., & Königl, A. 1979, ApJ, 232, 34
14. Blandford, R. D., & Levinson, A. 1995, ApJ, 441, 79
15. Begelman, M. C., Rees, M. J., & Sikora, M. 1994, ApJ, 429, L57
16. Bodo, G., Massaglia, S., Rossi, P., Rosner, R., Malagoli, A., & Ferrari, A. 1995, A&A, 303, 281
17. Bowman, M. 1994, MNRAS, 269, 137
18. Capetti, A., Axon, D. J., Kukula, M., Macchetto, F., Pedlar, A., Sparks, W. B., & Boksenberg, A. 1995, ApJ, 454, L85
19. Cawthorne, T. V. 1991, in Beams and Jets in Astrophysics, ed. P. A. Hughes (Cambridge: Cambridge Univ. Press), 187
20. Cawthorne, T. V., & Cobb, W. K. 1990, ApJ, 350, 536
21. Cawthorne, T. V., Wardle, J. F. C., Roberts, D. H., & Gabuzda, D. C. 1993, ApJ, 416, 519
22. Daly R. A., & Marscher A. P., 1988, ApJ, 334, 539
23. de Gouveia Dal Pino, E. M. 1999, ApJ, 526, 862
24. Dubal, M. R., & Pantano, O. 1993, MNRAS, 261, 203
25. Duncan, G. C., & Hughes, P. A. 1994, Apj, 436, L119
26. Eichler, D., & Wiita, P. J. 1978, Nature, 274, 38
27. Ferrari, A. 1998, ARA&A, 36, 539
28. Fosbury, R. A. E., Morganti, R., Wilson, W., Ekers, R. D., di Serego Alighieri, S., & Tadhunter, C. N. 1998, MNRAS, 296, 701
29. Gabuzda, D. C., & Cawthorne, T. V. 1992, in Variability of Blazards, ed. E. Valtaoja & M. Valtonen (Cambridge: Cambridge Univ. Press), 238
30. Gabuzda, D. C., Wardle, J. F. C., Roberts, D. H., Aller, M. F., & Aller M. H. 1994, ApJ, 435, 128
31. Gabuzda, D. C., Pushkarev, A. B., & Cawthorne, T. V. 1999, MNRAS, 307, 725
32. Gabuzda, D. C., Kochenov, P. Yu., Cawthorne, T. V., & Kollgaard, R. I. 2000, MNRAS, 313, 627
33. Gabuzda, D. C., Kochenov, P. Yu., Kollgaard, R. I., & Cawthorne, T. V. 2000, MNRAS, 315, 229

34. Gabuzda, D. C., & Cawthorne, T. V. 2000, MNRAS, 319, 1056
35. Gabuzda, D. C., Pushkarev, A. B., & Cawthorne, T. V. 2000, MNRAS, 319, 1109
36. Gabuzda, D. C., Kochenov, P. Yu., & Cawthorne, T. V. 2000, MNRAS, 319, 1125
37. Gabuzda, D. C., & Gómez, J. L. 2001, MNRAS, 320, L49
38. Gallant, Y. A., Achterberg, A., & Kirk, J. G. 1999, A&AS, 138, 549
39. Gallant, Y. 2001, these proceedings
40. Gallimore, J. F., Holloway, A. J., Pedlar, A., & Mundell, C. G. 1998, A&A, 333, 13
41. García-Miró, C., Gómez, J. L., & Alberdi, A. 1999, A&A, 351, 147
42. Giovanoni, P. M., & Kazanas, D. 1990, Nature, 345, 319
43. Gómez, J. L., Alberdi, A., & Marcaide, J. M. 1993, A&A, 274, 55-68
44. Gómez, J. L., Alberdi, A., & Marcaide, J. M. 1994, A&A, 284, 51
45. Gómez, J. L., Alberdi, A., Marcaide, J. M., Marscher, A. P., & Travis, J. P. 1994, A&A, 292, 33-44
46. Gómez, J. L., Martí, J. M., Marscher, A. P., Ibáñez, J. M., & Marcaide, J. M. 1995, ApJ, 449, L19
47. Gómez, J. L., Martí, J. M., Marscher, A. P., Ibáñez, J. M., & Marcaide, J. M. 1996, in ASP Conf. Ser. 100, Energy Transport in Radio Galaxies and Quasars, ed. P. E. Hardee, A. H. Bridle, & J. A. Zensus (San Francisco: ASP), p. 159
48. Gómez, J. L., Marscher, A. P., Martí, J. M., & Ibáñez, J. M. 1996, in Blazar Continuum Variability, ed. H. R. Miller, J. R. Webb, and J. C. Noble, p. 242
49. Gómez, J. L., Martí, J. M., Marscher, A. P., Ibáñez, J. M., & Alberdi, A. 1997, ApJ, 482, L33
50. Gómez, J. L., Marscher, A. P., Alberdi, A., Martí, J. M., & Ibáñez, J. M. 1998, ApJ, 499, 221
51. Gómez, J. L., Marscher, A. P., Alberdi, A., & Gabuzda, D. C. 1999, ApJ, 519, 642
52. Gómez, J. L., Marscher, A. P., & Alberdi, A. 1999, ApJ, 521, L29
53. Gómez, J. L., Marscher, A. P., & Alberdi, A. 1999, ApJ, 522, 74
54. Gómez, J. L., Marscher, A. P., Alberdi, A., Martí, J. M., Ibá nez, J. M., & Marchenko, S. G. 1999, in ASP Conf. Ser. 159, BL Lac Phenomenon, ed. L. O. Takalo, & A. Sillanpää (San Francisco ASP), 435
55. Gómez, J. L., & Marscher, A. P. 2000, ApJ, 530, 245
56. Gómez, J. L., Marscher, A. P., Alberdi, A., Jorstad, S. G., & García-Miró, C. 2000, Science, 289, 2317
57. Gómez, J. L. 2001, in Galactic Relativistic Jet Sources, ed. A. J. Castro-Tirado, J. Greiner, J. M. Paredes, Astrophysics and Space Science, in press
58. Hardee, P. E., Rose, A., Hughes, P. A., & Duncan, G. C. 1998, ApJ, 500, 599
59. Hardee, P. E. 2000, ApJ, 533, 176
60. Homan, D. C., & Wardle, J. F. C. 1999, AJ, 118, 1942
61. Hjellming, R. M., & Rupen, M. P. 1995, Nature, 375, 464
62. Hua, C. T. 1988, A&A, 199, 105
63. Hughes, P. A., Aller, H. D., & Aller, M. F. 1985, ApJ, 298, 301
64. Hughes, P. A., Aller, H. D., & Aller, M. F. 1989, ApJ, 341, 54
65. Hughes, P. A., Aller, H. D., & Aller, M. F. 1991, ApJ, 374, 57
66. Hughes, P. A., Duncan, C., & Mioduszewski, A. 1996, in ASP Conf. Ser. 100, Energy Transport in Radio Galaxies and Quasars, ed. P. E. Hardee, A. H. Bridle, & J. A. Zensus (San Francisco: ASP), p. 137
67. Hughes, P. A. 2000, in ASP Conf. Ser. 201, Particles and Fields in Radio Galaxies, ed. R. A. Laing, & K. M. Blundell (San Francisco: ASP). Available as astro-ph/0011127

68. Jones, T. W., & O'Dell, S. L. 1997, ApJ, 214, 522
69. Jones, T. W., & O'Dell, S. L. 1997, ApJ, 215, 236
70. Jones, T. W. 1988, ApJ, 332, 687
71. Jones, T. W., Ryu, D., & Engel, A. 1999, ApJ, 512, 105
72. Junor, W., & Biretta, J. A. 1995, AJ, 109, 500-506
73. Junor, W., Salter, D. J., Saikia, D. J., Mantovani, F., & Peck, A. B. 1999, MNRAS, 308, 955
74. Junor, W., Biretta, J. A., & Livio, M. 1999, Nature, 401, 891
75. Jorstad, S. G., Marscher, A. P., Mattox, J. R., Wehrle, A. E., Bloom, S. D., & Yurchenko, A. V. 2001, ApJS, in press
76. Kellermann, K. I., Vermeulen, R. C., Zensus, J. A., & Cohen, M. H. 1998, AJ, 115, 1295
77. Kirk, J. G., Rieger, F. M., & Mastichiadis, A. 1998, A&A, 333, 452
78. Kirk, J. G., Guthmann, A. W., Gallant, Y. A., & Achterberg, A. 2000, ApJ, 542, 235
79. Koide, S., Shibata, K., & Kudoh, T. 1998, ApJ, 495, L63
80. Koide, S., Shibata, K., & Kudoh, T. 1999, ApJ, 522, 727
81. Koide, S., Meier, D. L., Shibata, K., & Kudoh, T. 2000, ApJ, 536, 668
82. Komissarov, S. S., Falle, S. A. E. G. 1996, in Blazar Continuum Variability, ed. H. R. Miller, J. R. Webb, and J. C. Noble, p. 173
83. Komissarov, S. S., & Falle, S. A. E. G. 1997, MNRAS, 288, 833
84. Komissarov, S. S. 1999, MNRAS, 308, 1069
85. Königl, A. 1981, ApJ, 243, 700
86. Krichbaum, T. P., Hummel, C. A., Quirrenbach, A., Schalinski, C. A., & Witzel, A. 1990, A&A, 230, 271
87. Laing, R. A. 1996, in ASP Conf. Ser. 100, Energy Transport in Radio Galaxies and Quasars, ed. P. E. Hardee, A. H. Bridle, & J. A. Zensus (San Francisco: ASP), 241
88. Laing, R. A. Parma, P., de Ruiter, H. R., & Fanti, R. 1999, MNRAS, 306, 513
89. Lister, M. L., Marscher, A. P., & Gear, W. K. 1998, ApJ, 504, 702
90. Lister, M. L., & Smith, P. S. 2000, ApJ, 541, 66
91. Mannheim, K. 1993, A&A, 269, 67
92. Mantovani, F., Junor, W., Bondi, M., Cotton, W., Fanti, R., Padrielli, L., Nicolson, G. D., & Salerno, E. 1998, A&A, 332, 10
93. Marscher, A. P. 1980, ApJ, 235, 386
94. Marscher, A. P., & Gear, W. K. 1985, ApJ, 298, 114-127
95. Marscher, A. P., Gear, W. K., & Travis, J. P. 1992, in Variability of Blazars, ed. E. Valtaoja, & M. Valtonen (Cambridge: Cambridge Univ. Press), 85
96. Marscher, A. P. 1996, in ASP Conf. Ser. 100, Energy Transport in Radio Galaxies and Quasars, ed. P. E. Hardee, A. H. Bridle, & J. A. Zensus (San Francisco: ASP), 45
97. Marscher, A. P. 1996, in ASP Conf. Ser. 110, Blazar Continuum Variability, ed. H. R. Miller, J. R. Webb, & J. C. Noble (San Francisco: ASP), 248
98. Marscher, A. P. 1998, in ASP Conf. Ser. 144, Radio Emission from Galactic and Extragalactic Compact Sources, ed. J. A. Zensus, G. B. Taylor, & J. M. Wroble (San Francisco ASP), 25
99. Martí, J. M., Müller, E., & Ibáñez, J. M. 1994, A&A, 281, L9
100. Martí, J. M., Müller, E., Font, J. A., & Ibáñez, J. M. 1995, ApJ, 448, L105
101. Martí, J. M. 1997, in Relativistic Jets in AGNs, ed. M. Ostrowski, M. Sikora, G. Madejski & M. Begelman, Astronomical Observatory of the Jagiellonian University, p. 90

102. Martí J. M., Müller E., Font J. A., Ibáñez, J. M., Marquina A., 1997, ApJ, 479, 151
103. Martí, J. M., & Müller, E. 1999, Living Reviews in Relativity, http://www.livingreviews.org/Articles
104. Martí, J. M. 2000, in AGNs in their Cosmical Environment, eds. B. Rocca-Volmerange, & H. Sol, EDP Series in Astronomy and Astrophysics (EDP Sciences), in press
105. Meier, D. L. 2001, in Galactic Relativistic Jet Sources, ed. A. J. Castro-Tirado, J. Greiner, J. M. Paredes, Astrophysics and Space Science, in press
106. Meier, D. L., Koide, S., & Uchida, Y. 2001, Science, 291, 84
107. Mioduszewski, A. J., Hughes, P. A., & Duncan, G. C. 1997, ApJ, 476, 649
108. Mundell, C. G., Pedlar, A., Baum, S. A., O'Dea, C. P., Gallimore, J. F., & Brinks, E. 1995, MNRAS, 272, 355
109. Norman, M. L. 1996, in ASP Conf. Ser. 100, Energy Transport in Radio Galaxies and Quasars, ed. P. E. Hardee, A. H. Bridle, & J. A. Zensus (San Francisco: ASP), p. 405
110. Oosterloo, T. A., Morganti, R., Tzioumis, A., Reynolds, J., King, E., McCulloch, P., & Tsvetanov, Z. 2000, AJ, 119, 2085
111. Pacholczyk, A. G., 1970. Radio Astrophysics. Freeman, San Francisco
112. Peck, A. B., Taylor, G. B., & Conway, J. E. 1999, ApJ, 521, 103
113. Rayburn, D. R. 1977, MNRAS, 179, 603
114. Rickett, B. J., Quirrenbach, A., Wegner, R., Krichbaum, T. P., & Witzel, A. 1995, A&A, 293, 479
115. Rosen, A., Hughes, P. A., Duncan, G. C., & Hardee, P. E. 1999, ApJ, 516, 729
116. Rybicki, G., Lightman, A. 1979, Radiative Processes in Astrophysics. Wiley, New York
117. Salvati, M., Spada, M., & Pacini, F. 1998, ApJ, 495, L19
118. Seielstad, G. A., Cohen, M. H., Linfield, R. P., Moffet, A. T., Romney, J. D., Schilizzi, R. T., & Shaffer, D. B. 1979, ApJ, 229, 53
119. Steffen, W., Gómez, J. L., Williams, R. J. R., Raga, A. C., & Pedlar, A. 1997, MNRAS, 286, 1032
120. Steffen, W., Gómez, J. L., Raga, A. C. & Williams, R. J. R. 1997, ApJ, 491, L73
121. Swain, M. R., Bridle, A. H., & Baum, S. A. 1998, ApJ, 507, L29
122. Tadhunter, C. N., Villar-Martín, M., Morganti, R., Bland-Hawthorn, J., & Axon, D. 2000, MNRAS, 314, 849
123. Taylor, G. B. 1996, ApJ, 470, 394
124. Taylor, G. B. 1998, ApJ, 506, 637
125. Tingay, S. J. et al. 1995, Nature, 374, 14
126. Udomprasert, P. S., Taylor, G. B., Pearson, T. J., Roberts, D. H. 1997, ApJ, 483, L9
127. van Putten, M. H. P. M. 1993, ApJ, 408, L21
128. van Putten, M. H. P. M. 1996, ApJ, 467, L56
129. Villar-Martín, M., Tadhunter, C., Morganti, R., Clark, N., Killeen, N., & Axon, D. 1998, A&A, 332, 479
130. Wagner, S. J., & Witzel, A. 1995, ARA&A, 33, 163
131. Walker, R. C. 1997, ApJ, 488, 675
132. Wang, Z., Wiita, P. J., & Hooda, J. S. 2000, ApJ, 534, 201
133. Wardle, J. F. C., Homan, D. C., Ojha, R., & Roberts, D. H. 1998, Nature, 395, 457
134. Wehrle A. E. et al., 1997, in ASP Conf. Ser. 110, Blazar Continuum Variability, ed. H. R. Miller, J. R. Webb, & J. C. Noble (San Francisco: ASP), 430

135. Wilson, M. J., & Scheuer, P. A. G. 1983, MNRAS, 205, 449
136. Wilson, M. J. 1987, MNRAS, 226, 447
137. Xu, J., Hardee, P. E., & Stone, J. M. 2000, ApJ, 543, 161

3D Relativistic Hydrodynamics

Miguel A. Aloy[1] and José M. Martí[2]

[1] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
[2] Departamento de Astronomía y Astrofísica, Universidad de Valencia, C/ Doctor Moliner 50, 46100 Burjassot, Spain

Abstract. We review the evolution of the numerical techniques applied in relativistic hydrodynamics since the sixties until today. We focus our attention on the state–of–the–art high–resolution shock–capturing methods and the astrophysical applications involving three–dimensional simulations.

1 Introduction

A relativistic description of fluid dynamics is necessary in situations where the local velocity of the flow is close to the light speed in vacuum (c) or where the local internal energy density is comparable (or larger) than the local rest–mass density of the fluid. Alternatively, a relativistic description should be used whenever matter is influenced by large gravitational potentials ($\approx O(c^2) \approx 9\,10^{20}$ erg g^{-1}), where a description in terms of the Einstein field theory of gravity is necessary. Relativistic flows are present in numerous astrophysical phenomena, from stellar to galactic (and even cosmological –early Universe, galaxy formation–) scales. Among these phenomena are core collapse supernovae, X-ray binaries, pulsars, coalescing neutron stars and black holes, micro-quasars, active galactic nuclei, superluminal jets, gamma-ray bursts and, in general, any astrophysical scenario involving compact objects.

More than thirty years ago, the pioneering work of May and White [103], studying the process of stellar core collapse in spherical symmetry, triggered the use of relativistic numerical simulations as a tool to get insight into the above mentioned phenomena, complementing theoretical models and observations. However, the description of the ultrarelativistic regime has only been possible after the development, in the last ten years, of new numerical algorithms (the so called high–resolution shock–capturing, HRSC, techniques).

This work is aimed to summarize the basic theory of the numerical techniques applied to solve multidimensional (specifically three-dimensional, 3D) problems in the frame of relativistic hydrodynamics (RHD). Recent works reviewing this topic are [102] (in Special Relativity) and [45] (in General Relativity). In addition, we discusse the state of art of the current numerical 3D RHD codes. The work is organized as follows. The equations of General Relativistic Hydrodynamics (GRHD) are introduced in §2. Some important mathematical properties of the GRHD equations (conservative character, hyperbolicity) are discussed in Sections 2.1 and 2.2. Section 2.3 points out briefly the main differences between

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 197–226, 2002.

classical and relativistic equations. Sections 3, 4 and 5 form the body of the review. Section 3 is devoted to discuss different approaches for the integration of the RHD eqs. paying special attention to the most recent numerical algorithms. Present numerical 3D RHD codes are reviewed in §4 as well as some computational issues relevant for three dimensional simulations. Several astrophysics applications are discussed in §5. We finish this work with a summary (6).

2 The Equations of General Relativistic Hydrodynamics

The equations that describe the evolution of a relativistic fluid can be written as covariant divergences,

$$\nabla \cdot \mathbf{J} = 0 \,, \tag{1a}$$

$$\nabla \cdot \mathbf{T} = 0 \,, \tag{1b}$$

representing the conservation of rest-mass and energy–momentum in the space–time $\mathcal{M}$, described by a metric $\mathbf{g}$. In the previous equations, $\mathbf{J}$ is the current of rest mass and $\mathbf{T}$ is the energy-momentum tensor. For perfect fluids (*i.e.,* those without shear or heat conduction) and using a system of natural units ($G = c = 1$; G is the gravitational constant), the components of $\mathbf{J}$ and $\mathbf{T}$ on a coordinate basis are

$$J = \rho u^{\mu} \,, \qquad \mu, \nu = 0, \ldots, 3 \tag{2a}$$

$$T^{\mu\nu} = \rho h u^{\mu} u^{\mu} + p g^{\mu\nu} \,, \tag{2b}$$

ρ, p, h, ε and u^{μ} being the rest–mass density, the pressure, the specific enthalpy ($h = 1 + \varepsilon + p/\rho$), the specific internal energy and the four–velocity of the fluid, respectively. The system of equations (1a, 1b) is closed making use of the normalization condition of the four velocity ($u^{\mu} u_{\mu} = -1$, where summation is extended over repeated indices) and an equation of state (EoS), usually of the form $p = p(\rho, \varepsilon)$.

A consistent numerical simulation of the flow evolution within the framework of Relativity requires, in principle, the solution of the RHD equations (relativistic counterparts of the Euler or Navier–Stockes equations) coupled to the *full* set of Einstein field equations (that control the evolution of the space–time). However, the problem is so complex (especially in multidimensions), and the variety of astrophysical phenomena is so wide that for practical purposes it is worthy to consider different physical approximations. The easiest one is to consider that the gravitational field is unimportant and, therefore, the space–time is flat or, in other words, the metric of the space–time is the Minkowski metric. This is the approach of Special Relativistic hydrodynamic (SRHD) simulations and has been successfully applied to, *e.g.,* extragalactic jets (*e.g.,* [101]), afterglows of gamma–ray bursts (GRBs) [122], and also in other fields of physics, like *e.g.,* relativistic heavy–ion collisions [150]. Another possibility is to assume that the simulated flow is a *test fluid* evolving in an external static field created

by a massive object. This is the approximation followed in most GRHD problems. The GRHD approximation is fruitfully used in simulations of accretion of matter onto compact objects [63] and also in the context of the formation of GRBs [7] (see § 5.2). However, when strong and varying gravitational fields are encountered, the evolution of the metric has to be obtained by solving the Einstein field equations (EFE). This situation holds, *e.g.,* in the case of coalescing neutron stars (NSs) [86] or in the collapse of a massive star to a black hole (BH) – *e.g.,* [103] –.

2.1 Equations of GRHD in Conservation Form

Following the {3+1} formalism [9], and given a general coordinate system $\{x^\mu\} \equiv (t, x^i)$, the space–time $\mathcal{M}$ can be foliated into a set of spacelike hypersurfaces Σ_t (for each constant value of t) such that

$$\partial_t = \alpha \mathbf{n} + \beta^i \partial_i, \quad \mathbf{n} \cdot \partial_i = 0, \quad i = 1, 2, 3,$$

where $\{\partial_t, \partial_i\}$ define a coordinate basis and $\mathbf{n}$ is the unit timelike vector field normal to Σ_t.

The line element on $\mathcal{M}$ may be written in terms of the scalar *lapse*, α, the *shift* vector, β^i, and the three–metric of each hypersurface Σ_t, γ_{ij},

$$ds^2 = -(\alpha^2 - \beta_i \beta^i) dt^2 + 2\beta_i dx^i dt + \gamma_{ij} dx^i dx^j .$$

In coordinates $\{x^\mu\}$, equations (1a, 1b) are written (in conservation form) as

$$\frac{1}{\sqrt{-g}} \left[\frac{\partial \sqrt{\gamma} \mathbf{U}(\mathbf{W})}{\partial t} + \frac{\partial \sqrt{-g} \mathbf{F}^{(i)}(\mathbf{W})}{\partial x^i} \right] = \mathbf{S}(\mathbf{W}), \tag{3}$$

where g stands for the determinant of the four–metric of $\mathcal{M}$ and γ for the determinant of γ_{ij}, and where $\mathbf{U} \equiv (D, S_j, \tau)$ is the vector of unknowns, which are the rest-mass, momentum and energy densities, respectively, as measured by Eulerian observers (i.e., those having $\mathbf{n}$ as four–velocity). In the previous equations $\mathbf{F}^{(i)}$ are the vectors of fluxes

$$\mathbf{F}^{(i)}(\mathbf{W}) = \left[D \left(v^i - \frac{\beta^i}{\alpha} \right), S_j \left(v^i - \frac{\beta^i}{\alpha} \right) + p \delta^i_j, \tau \left(v^i - \frac{\beta^i}{\alpha} \right) + p v^i \right]$$

and $\mathbf{S}$ is the vector of sources

$$\mathbf{S}(\mathbf{W}) = \left[0,\ T^{\mu\nu} \left(\frac{\partial g_{\nu j}}{\partial x^\mu} - \Gamma^\delta_{\nu\mu} g_{\delta j} \right),\ \alpha \left(T^{\mu 0} \frac{\partial \ln \alpha}{\partial x^\mu} - T^{\mu\nu} \Gamma^0_{\nu\mu} \right) \right] .$$

Conserved variables are related with physical or *primitive* variables, $\mathbf{W} \equiv (\rho, \varepsilon, v^i)$, i.e., the rest-mass density and specific internal energy in the fluid rest frame and the fluid three–velocity measured by Eulerian observers, respectively, through

$$D = \rho \Gamma \tag{4a}$$

$$S_j = \rho h \Gamma^2 v_j \quad (j = 1, 2, 3) \tag{4b}$$

$$\tau = \rho h \Gamma^2 - p - \rho \Gamma, \tag{4c}$$

where

$$v_i = \frac{\mathbf{u}\cdot\partial_i}{-\mathbf{u}\cdot\mathbf{n}}$$
$$v^i = \gamma^{ij}v_j = \frac{u^i}{\alpha u^t} + \frac{\beta^i}{\alpha}\,,$$

and $\Gamma \equiv -\mathbf{u}\cdot\mathbf{n} = (1-\gamma_{ij}v^iv^j)^{-1/2}$ is the Lorentz factor.

2.2 GRHD Equations as a Hyperbolic System of Conservation Laws

Following, *e.g.*, [157], a system of q equations written in *conservation form*, like (3), is said to be hyperbolic at a point (t, x^i) if the Jacobian matrices of $\mathbf{F}^{(i)}$ ($\mathcal{B}^{(i)} = \partial\mathbf{F}^{(i)}/\partial\mathbf{U}$) have p real eigenvalues $\lambda_1, \ldots, \lambda_q$ and corresponding sets of p linearly independent right eigenvectors $\mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(q)}$.

The characteristic curves associated to the system are (in the one–dimensional case) the integral curves of

$$\frac{dx}{dt} = \lambda_k(\mathbf{U}(x,t)),\ \ k = 1, \ldots q \tag{5}$$

where we can notice that the eigenvalues of $\mathcal{B}^{(i)}$ are the velocities of propagation of the characteristic fields (characteristic speeds). Along each characteristic curve, there is one characteristic variable, *i.e.*, a component of $\mathcal{U}^{(i)} = \mathbf{L}^{(i)}\mathbf{U}$ ($\mathbf{L}^{(i)}$ is the matrix of left eigenvectors of $\mathcal{B}^{(i)}$), which remains constant. This fact allows one to construct formally solutions at advanced times by propagating the values of characteristic variables at previous times along the characteristic curves, i.e., by solving initial value problems (IVPs).

Anile [8] has shown that system (3) is hyperbolic for causal EoS, *i.e.*, for those where the local sound speed, c_s, defined by

$$hc_s^2 = \frac{\partial p}{\partial \rho} + (p/\rho^2)\frac{\partial p}{\partial \epsilon}, \tag{6}$$

satisfies $c_s < 1$.

In the particular case of SRHD the set of eigenvalues associated to the Jacobian matrix in the i–direction are [43,35,3]

$$\lambda_0 = v^i, \tag{7a}$$
$$\lambda_\pm = \frac{1}{1-v^2c_s^2}\left(v^i(1-c_s^2) \pm c_s\sqrt{(1-v^2)[1-v^2c_s^2-v^iv^i(1-c_s^2)]}\right). \tag{7b}$$

Let us notice that: (i) there exists a strong coupling between the components of velocity along the different spatial directions through the modulus of the velocity, v; (ii) in the one–dimensional case (1D) the expressions of $\lambda_\pm$ (associated to the acoustic waves) reduce to the Lorentz addition of the flow velocity and the local sound speed

$$\lambda_\pm = \frac{v \pm c_s}{1 \pm vc_s}; \tag{8}$$

(iii) $\lambda_\pm \to v \pm c_s$ in the Newtonian limit ($v, c_s \to 0$), and $\lambda_\pm \to 1$ in the ultrarelativistic limit ($v \to 1$).

For the sake of completeness let us say that the eigenvalues and right eigenvectors corresponding to the Jacobian matrices of fluxes in system (3) appear explicitly in ref. [12]. The explicit expression of the left eigenvectors can be found in ref. [3].

An important property of hyperbolic systems of conservation laws is that they admit discontinuous solutions (shocks). These discontinuous solutions satisfy the Rankine-Hugoniot (RH) jump conditions which establish the continuity of mass, momentum and energy fluxes across shocks. In the case of SRHD these conditions [154] read:

$$[\rho u^\mu] r_\mu = 0 \,, \tag{9}$$

$$[T^{\mu\nu}] r_\mu = 0 \,, \tag{10}$$

r_μ being the unit normal to the hypersurface of the space–time, σ, containing the discontinuity. The square brackets define the jump of a given variable across σ ($[F] = F_a - F_b$; where F_a, F_b are the values of F on the two sides of σ).

Considering σ as normal to the x-axis, a suitable choice of r_ν is $r^\nu = \Gamma_s(V_s, 1, 0, 0)$, where V_s is interpreted as the coordinate velocity of the hypersurface that defines the position of the shock wave and $\Gamma_s = (1 - V_s^2)^{-1/2}$ is the Lorentz factor associated to the shock.

RH conditions (9), (10) can be written in terms of the conserved and primitive quantities and the invariant mass flux across the shock, j ($j \equiv \Gamma_s D_a \{V_s - v_a^x\} = \Gamma_s D_b \{V_s - v_b^x\}$),

$$[v^x] = -\frac{j}{\Gamma_s} \left[\frac{1}{D}\right], \tag{11}$$

$$[p] = \frac{j}{\Gamma_s} \left[\frac{S^x}{D}\right], \tag{12}$$

$$\left[\frac{S^{y,z}}{D}\right] = 0, \text{ or } [h\Gamma v^{y,z}] = 0, \tag{13}$$

$$[v^x p] = \frac{j}{\Gamma_s} \left[\frac{\tau}{D}\right]. \tag{14}$$

Equation (13) implies that the orientation of the tangential velocity does not change across shocks. This also holds for selfsimilar expansions or rarefaction waves. However, the values of the components of the tangential velocity, (v^y, v^z), may jump across the shock (contrary to Newtonian hydrodynamics). In the case of contact discontinuities (no mass flux across the discontinuity, $j = 0$) arbitrary jumps in the tangential velocity are allowed although pressure and normal velocity should be continuous.

2.3 Newtonian versus Relativistic Hydrodynamics

The classical Euler equations are easily recovered from the RHD equations in the limit $c \to \infty$. The limits of the conserved variables and flux vectors lead to the corresponding quantities in the classical case:

$$\mathbf{U} = (D, S^j, \tau) \qquad \to (\rho, \rho v^j, \frac{1}{2}\rho \mathrm{v}^2 + \rho\varepsilon)$$

$$\mathbf{F}^{(i)} = (Dv^i, S^j v^i + p\delta^{ji}, S^i - Dv^i) \to (\rho v^i, \rho v^j v^i + p\delta^{ji}, v^i(\frac{1}{2}\rho \mathrm{v}^2 + \rho\varepsilon + p))$$

$$i, j = 1, 2, 3$$

The equations keep their conservative and hyperbolic characters but there are several factors that make RHD more complex to solve numerically than classical hydrodynamics: (i) the RHD equations are tightly coupled through Γ and h and, therefore, they display a larger non-linearity; (ii) there is no explicit relation between $\mathbf{W}$ and $\mathbf{U}$ (except for particular EoS), *i.e.,* obtaining the primitive from the conserved variables needs an iterative numerical method; (iii) the tangential flow velocity can change across discontinuities (see § 2.2) and, in addition, the characteristic speeds may suffer from aberration (see the coupling between different directions in expression (7b)); (iv) in the ultrarelativistic limit ($v \to 1$), the eigenfields are degenerate ($\lambda_0 \to \lambda_\pm \to 1$) which triggers the appearance of very thin structures in the flow (like, *e.g.,* in the case of relativistic blast waves) that may become a source of numerical errors; and (v) relativistic strong shocks can display unbounded jumps in physical variables; *e.g.,* for an ideal gas (*i.e.,* with an EoS: $p = (\gamma - 1)\rho\varepsilon$), the compression ratio between the post– and pre–shocked densities of a relativistic strong shock is such that (see, *e.g.,* [155]) $\rho_b/\rho_a \leq (\gamma\Gamma_a + 1)/(\gamma - 1)$ which tends to infinity in the ultrarelativistic limit ($v_a \to 1$). This should be compared with the compression ratio in a Newtonian strong shock: $\rho_b/\rho_a \leq (\gamma + 1)/(\gamma - 1)$ which is $\sim 4 - 7$ for typical values of γ.

3 Numerical Integration of the RHD Equations

The first attempts to solve numerically the equations of RHD started in the late 60's [103]. Since then and up to the 80's the field has been dominated by the so called artificial viscosity (AV) techniques. This was the result of the application of von Neumann & Richtmyer's [162] ideas (in the framework of classical hydrodynamics) to RHD. Basically, the algorithms consisted of standard finite difference techniques together with *viscous* terms added to the equations to damp spurious oscillations near shocks. The AV terms were *non–consistent* because they were not based on the energy–momentum tensor of a viscous fluid. However, they provide an artificial dissipative mechanism that makes the shock transition smooth, *i.e.,* extended over several numerical zones. The approach needs a large dissipation (the amount of viscosity being problem dependent) to handle strong relativistic shocks and, therefore, AV schemes may be very diffusive (smearing out every discontinuity in the flow). In addition, the algorithms were not

conservative and, hence, did not guarantee the propagation of discontinuities at the correct (physical) velocities. Due to all these problems, traditional AV methods are not optimal both in the mildly and in the ultrarelativistic regimes (*i.e.,* $\Gamma \geq 2$).

Over and above all these considerations, it remains true that AV methods have been widely used in many fields of astrophysics and we summarize briefly some of their key contributions. The first 1D, Lagrangian, full GRHD code (*i.e.,* also evolving the EFE) corresponds to May & White [103] and it was applied to stellar core collapse. However, due to the Lagrangian character of the code, it was impossible to extend it to multidimensions. Another important improvement was made by Wilson during the 70's [164,165]. In ref. [164] he presented the first multidimensional (2D), Eulerian, full GRHD numerical code. This code (and some different versions of it) has been successfully applied to axisymmetric stellar collapse (*e.g.,* [127,151,148,114,111,40]), accretion onto compact objects (*e.g.,* [63,125]) and numerical cosmology (*e.g.,* [24]).

As an example of the performance of the methods, we reproduce in Fig. 1b the plot given by Norman & Winkler [118] showing the evolution of the relative errors of Centrella & Wilson's algorithm [24] for the case of the reflection of a mildly relativistic shock against a wall (see Fig. 1a for a schematic representation). The solution develops a shock that moves away from the wall. In the postshock state the gas is at rest and much hotter than the gas in the preshocked medium. It is noticeable from Fig. 1b that errors are larger than 5% even for values of Γ as small as $\simeq 2.3$.

In the mid eighties, Norman & Winkler [118] proposed a reformulation of the difference equations with artificial viscosity consistent with the relativistic dynamics of non-perfect fluids. Accurate results across strong relativistic shocks with large Lorentz factors were obtained in combination with adaptive mesh

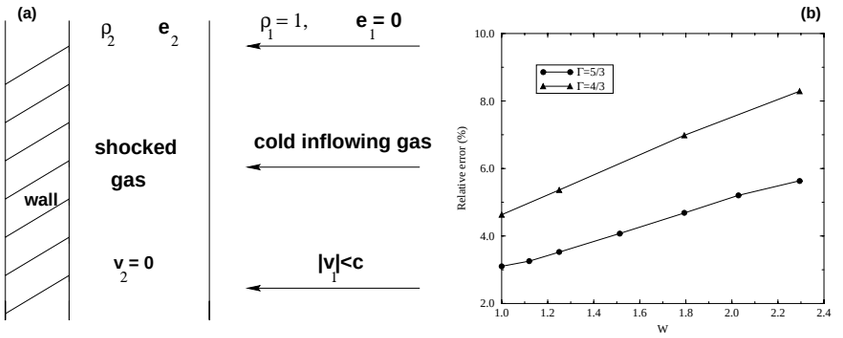

Fig. 1. (**a**) Scheme of the relativistic shock reflection test. (**b**) Relative errors of the relativistic shock reflection test as a function of the Lorentz factor (W in the plot) of the inflowing gas for two different values of the adiabatic index (Γ in the plot legends) using the explicit Eulerian techniques of [24]. Data from Centrella and Wilson [24]. Plot reproduced from Norman and Winkler [118].

techniques. However, the strong coupling introduced in the equations by the presence of the viscous terms in the definition of relativistic momentum and total energy densities required an implicit treatment of the difference equations and prevented the extension of the algorithm to multidimensions.

During the 1990's, a major break-through in the simulation of ultrarelativistic flows was accomplished when high–resolution shock–capturing (HRSC) methods, specially designed to solve hyperbolic systems of conservation laws, were applied to solve the RHD equations. Their application has caused a revolution in numerical RHD because (i) writing the equations in conservation form, guarantees convergence to the physically correct solution; (ii) HRSC methods exploit the hyperbolic character of the RHD equations and, therefore, the solution automatically satisfies RH conditions (this is the reason why they are called *shock–capturing*); and (iii) they provide high resolution in the sense that they have high order of accuracy in smooth regions of the flow while keeping discontinuities stable and sharp.

3.1 Basic Procedure of HRSC Methods

As in any finite difference or finite volume scheme, the first step consists in discretizing the equations on a finite numerical grid (t^n, x_j)

$$x_j = (j - 1/2)\Delta x, \quad j = 1, 2, \dots , \tag{15}$$

$$t^n = n\Delta t, \quad n = 0, 1, 2, \dots , \tag{16}$$

Δt, Δx being the time step and the zone size, respectively. The discretization of the system is such that the time evolution of zone averaged state vectors, $\mathbf{U}_j^n$, is governed by functions called *numerical fluxes*, $\hat{\mathbf{F}}_{j\pm1/2}$, evaluated at zone interfaces:

$$\frac{d\mathbf{U}_j^n}{dt} = -\frac{1}{\Delta x}\left(\hat{\mathbf{F}}_{j+1/2}(\mathbf{U}_{j-r}^n, \mathbf{U}_{j-r+1}^n, \dots, \mathbf{U}_{j+q}^n) - \hat{\mathbf{F}}_{j-1/2}(\mathbf{U}_{j-r-1}^n, \mathbf{U}_{j-r}^n, \dots, \mathbf{U}_{j+q-1}^n)\right).$$

where p and q are positive integers. Usually, $\mathbf{U}_j^n$ is an approximation to the zone average of $\mathbf{U}$ defined by

$$\mathbf{U}_j^n = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{U}(t^n, x)dx\,, \tag{17}$$

which is consistent with the integral form of the system of conservation laws and, $\hat{\mathbf{F}}_{j\pm1/2}$ are the time–averaged fluxes across the interfaces $x_{j\pm1/2}$ ($x_{j\pm1/2} = (x_j + x_{j\pm1})/2$) :

$$\hat{\mathbf{F}}_{j\pm1/2} \approx \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \mathbf{F}(\mathbf{U}(t, x_{j\pm1/2}))dt\,. \tag{18}$$

Computation of Numerical Fluxes. Numerical fluxes must be consistent with the actual fluxes, *i.e.,* $\hat{\mathbf{F}}(\mathbf{U},\mathbf{U},\ldots,\mathbf{U}) = \mathbf{F}(\mathbf{U})$. In (18) the only unknown is the value of the solution at every cell interface, $\mathbf{U}(t, x_{j\pm1/2})$ and its dependence on time. There are several ways of calculating (18), that give rise to different HRSC schemes. One of the most fruitful subset of methods are called *Godunov–type methods* (because they were first used in Godunov [51]). The discretization procedure approximates the solution by a piecewise continuous function at each time step (see Fig. 2 bottom panel) and, therefore, at every cell interface a jump between adjacent zone averaged values may occur. The situation is such that locally (at each interface) at every time level, t^n, one has an IVP with piecewise constant initial data for the RHD equations, *i.e.,* a Riemann problem (RP; see, *e.g.,* [157]). In Godunov-type methods, $\mathbf{U}(t, x_{j\pm1/2})$ is calculated by solving these RPs at every zone interface. The exact solution of the RP for the Euler equations is known (*e.g.,* [30] and references therein) and consists of a set of constant states separated by centered rarefactions, contact discontinuities and/or shocks connecting the piecewise constant *left* and *right* states. The top panel of Fig. 2 shows a space–time diagram of the evolution of the solution

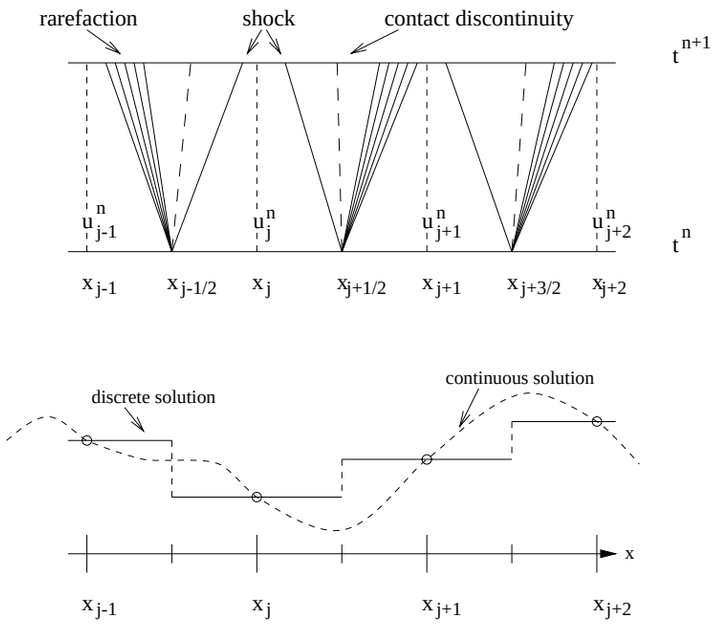

Fig. 2. Godunov's scheme: local solutions of Riemann problems. At every interface, $x_{j-\frac{1}{2}}$, $x_{j+\frac{1}{2}}$ and $x_{j+\frac{3}{2}}$, a local Riemann problem is set up as a result of the discretization process (bottom panel), when approximating the numerical solution by piecewise constant data. At time t^n these discontinuities decay into three elementary waves which propagate the solution forward to the next time level t^{n+1} (top panel). The time step of the numerical scheme must satisfy the Courant-Friedrichs-Lewy condition, being small enough to prevent the waves from advancing more than $\Delta x/2$ in Δt.

at several adjacent cell interfaces. The time step between two consecutive time levels, t^n, t^{n+1} must be limited in order to prevent the interaction of the solution of adjacent interfaces. This restriction on dt is known as Courant–Friedrich–Lewi (CFL) condition.

High Order of Accuracy. In order to understand the meaning of *high resolution* we need to introduce some mathematical concepts (see, *e.g.,* [87] for details). We say that the solution converges under grid refinement when the global error $||E_{\Delta x}||$, defined as

$$||E_{\Delta x}|| = \Delta x \sum_j |\bar{\mathbf{U}}_j^n - \mathbf{U}_j^n|, \tag{19}$$

tends to zero as $\Delta x \to 0$ ($\bar{\mathbf{U}}_j^n$ is the average of the *true* solution in the cell j). The idea behind definition (19) is that the finer the grid is, the better the numerical solution. Lax equivalence theorem [133] asserts that stability is a necessary condition to guarantee convergence. A measure of the stability of a solution can be its total variation at $t = t^n$, $\mathrm{TV}(\mathbf{U}^n)$, defined as

$$\mathrm{TV}(\mathbf{U}^n) = \sum_{j=0}^{+\infty} |\mathbf{U}_{j+1}^n - \mathbf{U}_j^n| \,. \tag{20}$$

A numerical scheme is said to be TV–stable, if $\mathrm{TV}(\mathbf{U}^{n+1}) \leq \mathrm{TV}(\mathbf{U}^n)$, $\forall n$ for any initial data. For non–linear, scalar conservation laws, a numerical scheme converges if it is written in conservation form with consistent numerical flux functions, and if it is TV–stable [87].

High–order accuracy in smooth parts of the flow is achieved by using conservative monotonic polynomial functions to interpolate the approximate solution within numerical cells. The reason to choose monotonic functions is that they lead to a decrease of the total variation of the solution (total–variation–diminishing schemes, TVD; [61]), ensuring stability. This interpolation or cell reconstruction provides more accurate left and right states for the Riemann problem by substituting the mean values $\mathbf{U}_j^n$ (that produces only first–order accuracy) by better representations of the true flow near the interfaces, thereupon, new RPs are setup at each cell interface. Different choices of the interpolation polynomial lead to TVD schemes of different order of accuracy. Piecewise constant functions provide first order accuracy (this was the original method of Godunov [51]). Second order of accuracy is obtained if piecewise linear functions are used (*e.g., monotonic upstream scheme for conservation laws* – MUSCL – method [158]) and, piecewise parabolic functions provide third order accuracy (*piecewise piecewise method* – PPM – [28]). Albeit it is possible to construct higher order polynomials to obtain better accuracy in smooth parts of the flow, it should be remarked that TVD methods degenerate to first order accuracy at extrema [119]. Therefore, other types of reconstruction (not being TVD) have been developed like, *e.g.,* total –variation–bounded (TVB) schemes [146], the *essentially non–oscillatory* (ENO) schemes [62] and the *piecewise–hyperbolic method*

(PHM) [93]. In other approaches, like the FCT algorithm (see, *e.g.*, [20]), a high order of accuracy is obtained by adding an anti–diffusive flux to a first order accurate numerical flux.

Time Advance. There are two main procedures to advance the conserved variables in time. One possibility is a standard discretization of the time derivative in system (17). In this case, the accuracy of the method depends on the order of accuracy up to which the numerical fluxes have been computed. The second alternative (known as the method of lines) regards (17) as a semidiscrete (*i.e.*, spatially discretized) system of ordinary differential equations to which any standard ordinary differential equation solver (*e.g.*, a Runge Kutta solver) can be applied.

3.2 Riemann Solvers

The key ingredient of Godunov–type methods is the solution of the RP at zone interfaces. Nevertheless, there is no exact closed–form to the RP neither for the RHD equations nor for the classical Euler equations, not even for ideal gases. Iterative algorithms are devised whereby the solution can be computed numerically up to any practical degree of accuracy. Numerical schemes aimed to solve RPs are known as Riemann solvers.

The solution of the RP can be either exact or approximate and, therefore, the Riemann solvers can be, at first instance, classified as exact or approximate Riemann solvers. Although the exact solution to the RP is known, there are a number of reasons to compute approximate solutions. The first is that the precision of any difference method is finite and, actually depends on the order of the method. Hence, an approximate RP solution with an accuracy better than that of the finite difference scheme might be acceptable. The second reason is that exact Riemann solvers are computationally expensive algorithms while approximate ones are more efficient. That is particularly important in multidimensional calculations because then numerical efficiency is a main issue. Finally, the solution of the RP is averaged in time to compute the numerical fluxes (see 18) and, therefore, the fine details of the solution are lost. Hence, it may pay off to evaluate an approximate (but accurate enough) solution to the RP.

Among the class of exact Riemann solvers, Godunov [51] is credited with the first exact Riemann solver for the Euler equations. Martí & Müller [98] extended this classical solution of the RP to the SRHD equations when the velocities parallel to the initial discontinuity (tangential velocities) are zero. Pons, *et al.* [130] have given the general solution (for non–vanishing tangential velocities) for the RP in SRHD. Wen, Panaitescu & Laguna [163] have extended Glimm's method or Random Choice Method [50,25] to 1D SRHD. Balsara [11] and Dai & Woodward [31] have extended to SRHD the two–shock approximation method of Colella [27] for classical fluid dynamics.

The strategy to construct approximate Riemann solvers (in both classical and relativistic hydrodynamics) relies on the local linearization of the Jacobian

matrices of the flux vectors, $\mathcal{B}^{(i)}$, extending a procedure introduced by Roe [132] and sometimes referred to as *local characteristic approach* (LCA). The idea of the LCA is to use the spectral decomposition of $\mathcal{B}^{(i)}$ to rewrite the original system as a new one of uncoupled scalar equations, in terms of the characteristic variables. The locally linear system can then be easily solved to obtain appropriate numerical fluxes for the original system. Sometimes the linearization process involves an averaged intermediate state at zone interfaces. This is the case of the original Roe solver [132] and its relativistic extension [39], and Martí *et al.* [97] or Falle & Komissarov [41] approaches. Donat & Marquina [36] have extended a numerical flux formula which was first proposed by Shu & Osher [147] for scalar equations to systems. In the scalar case and for characteristic wave speeds which do not change sign zone interfaces, Marquina's flux formula is identical to Roe's flux. Otherwise, the scheme switches to the more viscous, entropy satisfying local Lax–Friedrichs scheme [147]. In the case of systems, the combination of Roe and local–Lax–Friedrichs solvers is carried out in each characteristic field after the local linearization and decoupling of the system of equations [36]. However, contrary to Roe's and the previously cited linearized methods, the extension of Marquina's method to systems does not require on any averaged intermediate state. Marquina's flux formula has been successfully used in the ultrarelativistic regime and in 2D axisymmetric [100,101] and 3D SRHD [3] and even in 3D GRHD [7,44].

Finally, a very simple approach is an extension of the Harten, Lax, van Leer (HLL) solver to SRHD [139]. This method avoids the explicit calculation of the eigenvalues and eigenvectors of the Jacobian matrices and is based on an approximate solution of the original Riemann problems with a single intermediate state, $\mathbf{U}_*$, determined by requiring consistency of the approximate Riemann solution with the integral form of the conservation laws in a grid zone. The algorithm needs estimates of lower and upper bounds for the smallest and largest signal velocities, $a_{\rm L}$ and $a_{\rm R}$, respectively. Good estimates for $a_{\rm L}$ and $a_{\rm R}$ are essential to guarantee robustness and, at the same time, the minimal amount of numerical viscosity (the larger the difference $||a_{\rm L}| - |a_{\rm R}||$ is, the larger the viscosity of the method; if the difference is too small, undesired numerical oscillations around discontinuities appear). In the non-relativistic case, Einfeldt [37] proposed to use the smallest and largest eigenvalues of Roe's matrix (this corresponds to the HLLE solver). Duncan & Hughes [33] have generalized the method to 2D SRHD.

3.3 Other HRSC Schemes

Symmetric TVD Schemes with Nonlinear Numerical Dissipation. Symmetric TVD (sTVD) methods [32] are another subset of the HRSC methods. As Godunov–type methods, they are written in conservation form, but they are not based on solving Riemann problems. sTVD methods use standard finite difference methods (*e.g.,* Lax Wendroff scheme) and employ local conservative dissipation terms in order to stabilize the algorithm across discontinuities. The numerical dissipation term is local, free of problem dependent parameters and does not require any characteristic information. Hence, they are simpler than

Godunov–type schemes. Extensions to 2D and 3D general relativistic magneto hydrodynamics (GRMHD) can be found in, *e.g.*, [74,73,116].

Relativistic Beam Scheme. In the *beam scheme* [136] and its relativistic extension [167] the hydrodynamic equations are solved as the limit of the corresponding Botzmann equation. The velocity distribution functions are approximated by several Dirac delta functions or discrete beams of particles in each computational cell, which reproduce the appropriate moments of the distribution functions. This beam scheme, although being a particle method derived from a microscopic kinetic description, has all the desirable properties of modern characteristic–based wave propagation methods based on a macroscopic continuum description. Yang *et al.* [167] show that the integration scheme for the beams can be cast in the form of an upwind conservation scheme in terms of numerical fluxes and build up high–order variants of the relativistic code in terms of different TVD and ENO interpolations.

3.4 Other Approaches

Van Putten's Approach. Van Putten [159] solves the equations of (ideal) special relativistic magneto–hydrodynamics (SRMHD) formulating Maxwell's equations as a hyperbolic system in divergence form. State vectors and fluxes are decomposed into a spatially constant mean and a spatially dependent variational part. Then, the SRMHD equations become an evolution system for integrated (continuous) quantities for which standard integration methods can be used. In order to update the state vectors of the original system, a numerical differentiation of the integrated quantities is necessary. This process can lead to oscillations in the case of strong shocks and a smoothing algorithm needs to be supplied. The approach has been applied to SRMHD jets with values of the Lorentz factor up to 4.25 [160,161].

Relativistic Smoothed Particle Hydrodynamics. Smoothed particle hydrodynamics (SPH; [89]) represents a fluid by a Monte Carlo sampling of its mass elements. The motion and thermodynamics of these mass elements is governed by the hydrodynamic equations. As the equations involve continuous properties of the fluid, it is necessary to estimate these quantities from the positions, velocities and internal energies of the fluid mass elements, which are like particles moving with the flow. SPH treats the particle positions as a finite set of interpolating points where the continuous fluid variables and their gradients are estimated by an appropriately weighted average over neighboring particles. This means that it is a free–Lagrange method, *i.e.*, spatial gradients are evaluated without the use of a computational grid. As a result, the method provides a genuinely multidimensional (3D) description of the fluid with a lower computational cost (depending on the number of particles employed) than typical finite volume, HRSC, multidimensional methods. SPH has been extended to SRHD by Monaghan [109]. Other SRHD applications can be found in, *e.g.*, [84,91,92,26].

Implementations of SPH in GRHD may be seen in [70,83,145]. The artificial viscosity (necessary to handle discontinuities in the fluid) is the most critical issue for relativistic SPH codes. The reason is that, unlike its Newtonian analogue, the relativistic theory has not yet been developed to the degree required to predict – by a relativistic kinetic theory – the form of the dissipation terms. To overcome this problem, the most successful approach consists on taking the concepts from Riemann solvers as a guide to improve the artificial viscosity required in SPH (see [110,26]).

4 Computational Issues and Current 3D RHD Codes

4.1 Computational Issues

Due to the extreme computational resources demanded by typical 3D RHD simulations – as compared to 1D or 2D RHD ones –, technical improvements of the codes are required. Among the most important issues that should be addressed we find the optimization of the memory and the algorithm itself and, the design of suitable strategies for data analysis.

Memory Handling Optimization. Even small problems in 3D require huge amounts of RAM memory. As an example, 15 variables (e.g., a set of five conserved variables, five primitive variables and five fluxes) discretized in a volume containing 256^3 cells, in double precision ($\times 8$ bytes), would need ~ 2 Gb of RAM. Moreover, this number of 3D variables is a very moderate value that can easily be multiplied by a factor of ~ 4 if a GRHD problem with dynamical evolution of the space–time is considered, or if several chemical species are present in the fluid. Thereby, a reduction towards the minimal set of 3D variables is mandatory. In addition, the algorithms should be designed to allow for an efficient memory access (which is a computer dependent issue), the reason being that nowadays the speed of the memory systems is about one order of magnitude smaller than the actual CPU speeds. This last point is particularly critical when superscalar, cache–oriented computers (the most commonly used systems in supercomputing, at present) are used. Finally, high order methods are preferred in 3D, because less grid points are needed to resolve fine structures in the fluid (as compared with lower order methods).

Algorithm Optimization. A typical 3D RHD code, using an ideal EoS, may need about 10^{-4} s to update a numerical zone on a single processor. If one has to update 256^3 zones, the time per code iteration is roughly 30 minutes per processor. This means that 1000 time steps may need several days to a few weeks to be completed. This time estimate, can grow drastically if, *e.g.,* realistic microphysics and/or EoS are used. Hence, code efficiency becomes important in order to reduce the total computational time. An obvious way of decreasing the run time is to simplify analytically all the expressions, in particular those that may lead to numerical cancellation errors [5]. However, the largest impact

on performance comes from computer architecture considerations, *i.e.,* it is necessary to write algorithms oriented to the type of computer that is going to be used (cache based, vector based, etc.). Code parallelization and/or vectorization is unavoidable to work on current (both distributed or shared memory based) supercomputers. Finally, given that the affordable resolution in *fixed* 3D grids is still far from being appropriate to study many problems – even with the largest supercomputers –, resolution enhancement algorithms are crucial (like, *e.g.,* adaptive mesh refinement [15], multiresolution [60]).

Data Analysis. Applying the same data analysis than in 1D or 2D it is neither appropriate nor feasible in 3D, the reason being the huge data sets to be stored (an analyzed). A crude estimate of the typical storage requirements necessary to save every variable at every time step leads to unpractical (if not at all affordable) values of the order of terabytes of disc space. Albeit, it is convenient to design *a priori* the analysis strategy, *i.e.,* it is convenient to choose which variables have to be saved. An example of this kind of procedure is the synthesis of emission maps from simulations of relativistic jets including time–delay effects (see Sect. 5). In such a case, the values of the state variables in many different time levels are necessary in order to integrate the emission along rays parallel to the line of sight (for a given observing angle). In practice the number of time levels required to make the full integration is of the same order than the number of time steps of the simulation, which means that terabytes of data should be saved. An alternative is to choose *a priori* the viewing angles that are interesting and perform the integration on run time – picking out the values from the appropriate zones at each time level for every angle –.

4.2 Technical Status of the Existing 3D RHD Codes

There are, according to our knowledge, four different 3D RHD codes, used in astrophysical applications: Koide–Nishikawa's code [74,76], GENESIS [3], Cactus [44] and Shibata–Nakamura's code [143]. In the following we will describe some of the main features of each one.

The group of Koide and coworkers has developed a GRMHD code (for a fixed background metric) based on a sTVD scheme (see Sect. 3.3). The code has been applied to study the propagation of extragalactic jets through magnetized atmospheres [115] and also to simulate the early stages of jet formation by magnetohydrodynamic mechanisms in background BH spacetimes [75]. A series of tests of Koide et al.'s method involving mainly continuous solutions can be found in ref. [76]. Koide et al.'s code has proven to be very stable (although very difussive) when simulating mildly relativistic flows (maximum Lorentz factors ≈ 4) with discontinuities.

GENESIS [3] is a conservative 3D RHD code (used in astrophysical SRHD and GRHD applications) based on HRSC techniques. It uses Marquina's flux formula to compute numerical fluxes and a third order PPM spatial interpolation. The time advance is performed by means of a high–order Runge–Kutta method.

GENESIS has been extensively tested in problems involving strong shocks even in the ultrarelativistic regime, being able to handle Lorentz factors as large as $3 \cdot 10^5$ (in the one–dimensional wall reflection test). Simulations of extragalactic jets (using SRHD, at both parsec [6] and kiloparsec [4] scales) and progenitors of GRBs (in GRHD with a Schwarzschild background metric, [7]) have been performed with GENESIS.

Cactus is a numerical tool developed within a collaboration of the Numerical Relativity divisions at the National Center for Supercomputing Applications, the Albert Einstein Institut and the Washington University [22]. Cactus is able to solve the full set of Einstein field equations coupled to a perfect fluid source. The hydrodynamic evolution employs HRSC methods and can be computed by means of different Riemann solvers (Roe, Marquina) and flux–splitting schemes. The metric evolution may be followed with several formalisms (ADM, hyperbolic formulations, conformal–tracefree) combined with a number of integration methods (leapfrog, Crank–Nicholson, MacCormack along with Strang splitting) and various gauges (algebraic, maximal slicing, etc.). The main goal of the code is the simulation of astrophysical processes involving NSs and BHs. Font *et al.* [44] have tested the capabilities of the hydro part computing shock tubes. In Fig. 3, a comparison between Cactus and GENESIS is shown for a mildly relativistic Riemann problem. Despite the slightly different grids (100^3 for GENESIS, 128^3 for Cactus) the results are still comparable, the reason being that both codes use the same HRSC techniques. Font *et al.* [44] have also tested the GRHD coupling in Cactus simulating Friedman–Robertson–Walker cosmologies with dust and polytropic NSs (static and boosted). Alcubierre *et al.* [2] have performed a number of experiments (Brill waves, single BH, static boson and neutron stars) comparing different formulations of the Einstein equations. Recently, Landry & Teukolsky [86] have performed simulations of coalescing binary NSs. The sta-

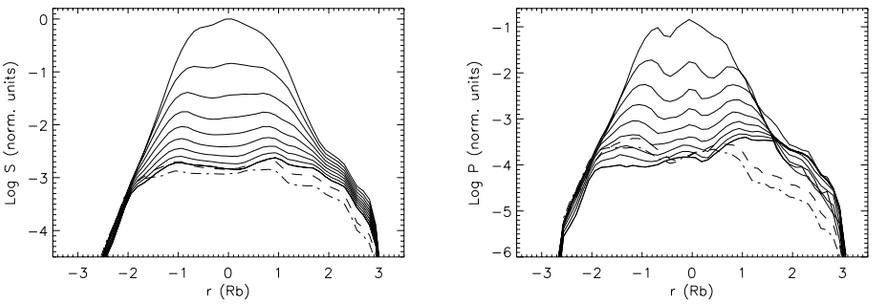

Fig. 3. Comparative performance of Cactus (left panel; figure from [44]) and GENESIS (right panel; figure from [3]) on a three–dimensional mildly relativistic shock tube test. The panels show several primitive variables along the main diagonal of the computational domain (solid line: analytic solution; symbols: numerical solution). The initial data are a constant left (L) and right (R) state characterized by $(\rho_L, p_L, v_L) = (10, 13.3, 0)$ and $(\rho_L, p_L, v_L) = (1, 0.66 \cdot 10^{-6}, 0)$. Numerical grid and evolution time is similar in both cases.

bility of the Einstein evolution is a main issue of this code, especially when the ADM or hyperbolic formulations are considered.

Shibata–Nakamura's code is based on the conformal–tracefree (CT) formulation of the Einstein equations [141,140,142]. The hydrodynamic equations are written (and solved) as a set of advection (i.e., non–conservative) equations to which AV is added as in Wilson's traditional approach (see § 3). The CT formulation is employed for the Einstein equations which are evolved by means of a leapfrog scheme. Several gauges are available (maximal slicing, minimum distortion). Shibata & Nakamura [140] have used this code to study the gravitational wave production of selected matter configurations. Rigidly rotating neutron star stability and neutron star mergers in 3D were computed in [113,143]. For the first time in 3D full GR, the dynamical collapse of a rotating neutron star has been simulated with this scheme [144]. Although the code –and the CT formulation– allows for the longest stable evolutions in dynamical space–times, the quality of the results, as measured by the degree of fulfillment of the constraint Einstein equations, is not as satisfactory as in the ADM or hyperbolic formulations. A comparison between the quality of the results of Shibata–Nakamura's code and GENESIS (using HRSC methods) is shown in Fig. 4 for the shock reflection test in a flat space–time. Shibata & Nakamura' results (for an inflow velocity as

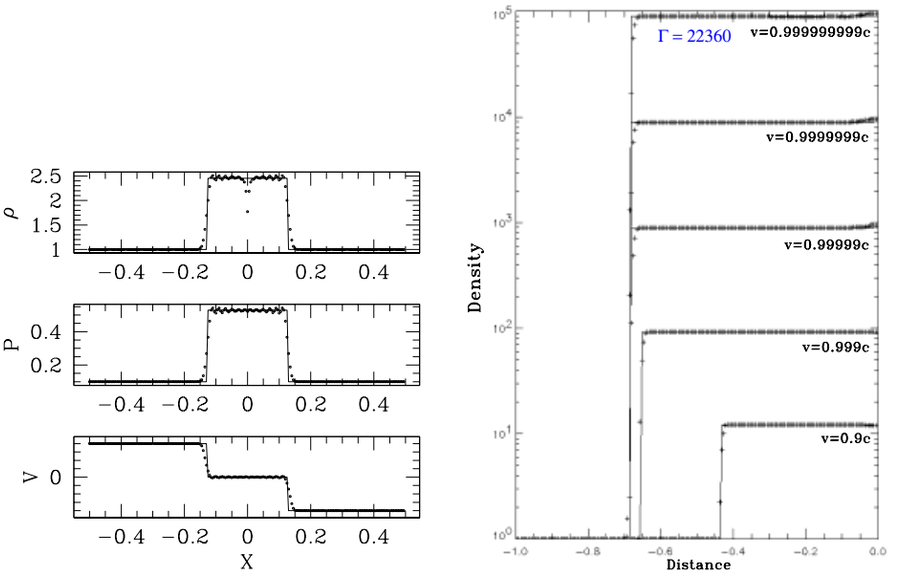

Fig. 4. Comparative performance of Shibata–Nakamura' code (left panel; figure from [143]) and GENESIS (right panel; figure from [3]) on the shock reflection test. In both cases the abscissas represent the distance to the reflection point. Left panel shows, from top to bottom, the density, pressure and flow velocity distributions, for an inflow velocity of $0.4c$. In the right panel, the density distribution (in logarithmic scale) is shown for a sample of inflow velocities –from $0.9c$ to $0.999999999c$–.

small as $0.4c$) display the well known pathologies of traditional relativistic AV methods in the treatment of shocks (diffusion, oscillations). HRSC techniques are much less diffusive (even with Lorentz factors as large as 22360) and produce more stable profiles in the postshock state.

5 Applications

In this Section we will review briefly some of the most relevant 3D applications of SRHD and SRMHD codes in astrophysical scenarios (i.e., relativistic jets) as well as recent GRHD and GRMHD simulations in the context of progenitors of GRBs and jet formation, respectively. Applications to dynamical space–times, in which Einstein equations are coupled to the GRHD equations, are beyond the scope of this article (for a review see, *e.g.*, [45]).

5.1 Relativistic Astrophysical Jets

In the standard model [19,138] the elongated radio structures connected to the center of AGNs in radio galaxies and radio–loud quasars are considered as continuous ejections of highly collimated, supersonic and very stable plasma. The emission is produced by synchrotron and inverse Compton processes of electrons accelerated up to ultrarelativistic energies in the vicinity of a *central engine*. The asymmetries in the radio flux of the two oppositely directed jets of a source and the superluminal motions observed in a few dozens of compact sources are explained by assuming that both jets propagate with relativistic speeds along directions at small angles to the line of sight. The relativistic Doppler beaming of the emission in the direction of motion accounts for the observed emission asymmetries whereas apparent superluminal speeds are explained by the combination of a finite value of the speed of light and the relativistic motion of the emitting source.

The formation, collimation and propagation of extragalactic jets involves scales ranging from some microparsecs to hundreds of kiloparsecs. Thereby, the study of the jet phenomenon is conveniently chopped into several pieces each one covering a smaller range of length scales. The mechanism governing the jet formation and collimation is still a challenge, mainly because the most detailed high–frequency VLBI observations of nearby radio sources can resolve at most the compact radio cores with a linear resolution of ≥ 0.1 pc [17][1], while the Schwarzschild radius ($R_s = GM/c^2$) of a $10^9 M_\odot$ galactic BH (the supposed central engine in the most commonly accepted interpretation) is $\sim 10^{-5}$ pc. Consistently, our theoretical view of jet formation is mainly constrained by the fact that many jets are well collimated by the time they have propagated to a distance $\leq$ 1pc from the nucleus (*e.g.*, [69]). Several mechanisms of jet formation have been proposed, all of which present some difficulties [13]. Additionally, there

[1] The VLBI Space Observatory Program (VSOP), allows for a linear resolution of $\sim 0.1 - 1$ pc even in distant sources like S5 0836+710 ($z = 2.17$; [88]).

is a wide variety in the observed properties of jets, so there may be a variety of jet collimation mechanisms. The standard model assumes that jet formation involves accretion onto a central compact object, such as a NS or a stellar BH in the case of galactic microquasars (GRS 1915+105 [107] and GRO J1655-40 [156]), or a supermassive BH in AGNs. This accretion fuels bipolar outflows which are further collimated and accelerated by MHD processes at subparsec scales (see, *e.g.*, [81]).

At parsec scales (from 0.1 pc to 100 pc) jets are observed using VLBI imaging. VLBI radio maps display highly collimated jets whose morphology is characterized by a bright spot at the jet end and a series of components which separate from the core, sometimes at superluminal speeds. Many parsec scale jets show intraday variability of the radio flux, excess in brightness temperatures and one-sidedness. The interpretation of these features is that the jet material moves at small angles to the line of sight with bulk Lorentz factors $\Gamma \simeq 15 h_{65}^{-1}$ [49] (h_{65} is Hubble's constant in units of 65 km/s Mpc^{-1}), or even larger ($\Gamma \simeq 30 - 100$) if the intraday variability is intrinsic and a result of incoherent synchrotron radiation [14]. The moving components are interpreted, within the *shock–in–jet* model as traveling shock waves [95,54]. A worthy byproduct of this model is the explanation of the complex multifrequency brightness and polarization variations in blazars.

At kiloparsec scales, sources are shaped by the interaction with the external medium, with a morphological dichotomy between FRI and FRII [42] sources whose basis seems to be the source power. The morphology of FR I sources is the result of a deceleration from relativistic to non–relativistic, transonic speeds [16,85], occurring within the first kpc. In the case of the FRIIs, mildly relativistic jet speeds (Lorentz factors between 2 and 3) well outside the galaxy have been inferred from flux asymmetries between jets and counter–jets [21] and superluminal motions at kpc scales (*e.g.*, in 1928+738 [65] and 1055+201, 1830+285, 2209+080 [64]). The mechanism by which the relativistic flows inferred from radio jets at pc scales remain collimated up to kpc scales is still unresolved.

Large-Scale Jets. At far enough distances from the central object (pc and kpc scales) the effects of gravity are negligible and we can assume (as a first approximation to the problem) that magnetic fields are dynamically unimportant. In this approach, numerical simulations using a pure SRHD treatment have been performed since the early 90s to study the morphology and dynamics of relativistic jets.

The development of codes based on HRSC techniques has allowed 2D axisymmetric time–dependent relativistic hydrodynamic simulations [33, 38, 99–101, 80, 134] to be performed. These simulations led to the conclusion that both the internal energy and the Lorentz factor of the beam enhance the stability of relativistic jets compared to their classical counterparts, through the increase of the effective inertial mass of the beam. Relativistic MHD simulations in 2D using pseudo–spectral techniques [159,161] or sTVD methods [74,73] have been another step forward in our understanding of relativistic astrophysical jets.

Only since 1998 the morphology and dynamics of relativistic jets is studied with 3D SRHD [3,4] or SRMHD [115,116] simulations. Aloy *et al.* [3] did a 3D simulation (in Cartesian coordinates) of an axisymmetric relativistic jet propagating through an homogeneous atmosphere. The simulated jet is characterized by a beam–to–external proper rest–mass density ratio $\eta = 0.01$, a beam Mach number $\mathcal{M}_b = 6.0$, and a beam flow speed $v_b = 0.99c$. An ideal gas equation of state with an adiabatic exponent $\gamma = 5/3$ is assumed to describe both the jet matter and the ambient gas. The beam is in pressure equilibrium with the ambient medium which fills a domain (X,Y,Z) with a size of $15R_b \times 15R_b \times 75R_b$ ($120 \times 120 \times 600$ cells), where R_b is the beam radius. The jet is injected at $z = 0$ in the direction of the positive z-axis through a circular nozzle. Simulations were typically performed with 16 R10000 processors (on a SGI-Origin 2000) and need about ten thousand time iterations. The execution time was about 100 hours. Genuine multidimensional effects were included by perturbing the axial injection velocity.

Koide and collaborators [115,116] simulated the evolution of relativistic jets through a magnetized uniform atmosphere during a very brief period of time. The numerical setup [116] consisted on a cylindrical jet injected through a circular nozzle into an oblique (45° with respect to the jet axis) magnetic atmosphere. The computational domain was a cubic box of $20R_b \times 20R_b \times 20R_b$ ($101 \times 101 \times 101$ cells). The jet had the following parameters: $v_b = 0.98756c$, ratio between the beam and magnetic specific energy densities $\varepsilon_b/\varepsilon_m = 1/3$, $\mathcal{M}_b = 4.0$, $\gamma = 5/3$ and considered two different strengths for the ambient magnetic field, *weak* and *strong*. The simulation lasted for 35 CPU hours on a SGI–Power Challenge and required about 1 Gb of RAM. It should be remarked that the coarse grid zoning used in Nishikawa's [116] simulations (5 cells/R_b), prevented them from studying genuine 3D effects in relativistic jets in detail.

Parsec Scale Jets. The presence of emitting flows at almost light speed enhances the importance of relativistic effects in the apparency of jets. This fact is stressed in the case of parsec scale jets, triggering the combination of synchrotron emission models and hydrodynamic simulations to compare with observations. Theoretical models aimed at explaining the appearance of radio jets [68,94,95,72] studied the spectrum of the synchrotron emission produced in different parts of the jet and the perturbations induced on the spectrum due to jet inhomogeneities. The inhomogeneities were introduced to mimic the ejection of components in real sources. The (frequency dependent) light curves of both total and polarized flux of the ejected components are understood as enhanced emission behind relativistic shock waves propagating down the underlying steady jet (*shock–in–jet* model).

The theoretical models gave rise to more detailed numerical models (see, *e.g.*, [67,96,53]) which, however, were restricted to a simplified (one-dimensional) hydrodynamic evolution. In 1995, Gómez *et al.* [55] produced the first numerical simulations of the pc scale synchrotron emission from 2D SRHD jets. Since then, other works have followed the same approach [56,34,106,79].

Very recently, Aloy *et al.* [6] have computed, for the first time, the emission from 3D relativistic jets. As in the previous 2D simulations, the approach followed by Aloy et al. rested on two main points. First the jet structure is calculated using a relativistic time–dependent hydrodynamic code. Second, the radio emission from the hydrodynamic jet models is computed by integrating the transfer equations of synchrotron radiation, once the appropriate relations between the computed hydrodynamic quantities and the emission/absorption coefficients are established. The procedure accounts for the appropriate opacity and relativistic effects, such as Doppler boosting and relativistic aberration. The simulations were done with the code GENESIS [3] and the radio emission was calculated with the same code as in Gómez *et al.* [55,56] (see also [58]). Aloy *et al.* [6] discuss some observational consequences of the interaction between the relativistic jet and the surrounding medium, leading to the development of a shear layer. The presence of such a layer (with distinct kinematic properties and magnetic field configuration) has been invoked *ad hoc* in the past by several authors [78,85] to account for a number of observational characteristics in FR I radio sources. However, its physical nature is still largely unknown. Remarkable effects associated with this layer are, *e.g.,* the presence of rails of low polarization intensity (see emission dips in Fig. 5–right) along the shear layer. Recently, Swain, Bridle, & Baum [152] have found evidence of such shear layers in FR II radio galaxies (3C353), and Attridge, Roberts, & Wardle [10] have inferred a two-component structure in the parsec scale jet of the source 1055+018.

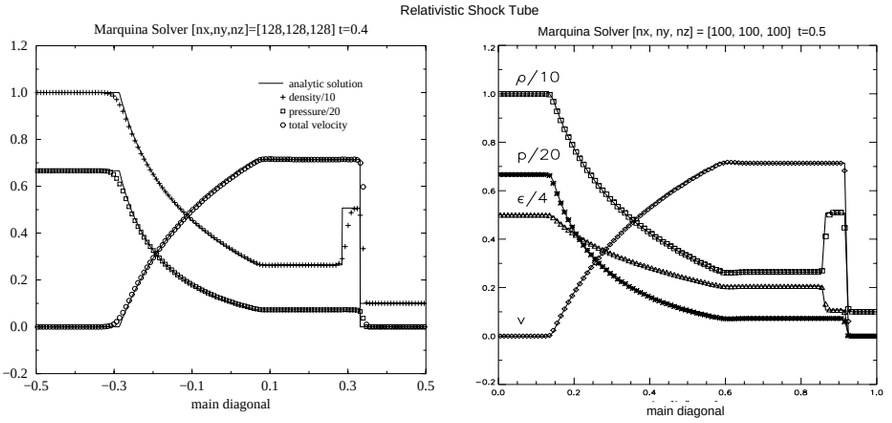

Fig. 5. Figure from [6]. Logarithm of the integrated total (*left*) and polarized (*right*) intensity across the jet for different viewing angles. Lines are plotted in intervals of 10° from an angle of 10° (top line in both plots), to 90° (showing a progressive decrease in emission). Dashed lines (dot dashed) correspond to an observing angle of -100° (-140°). Units are normalized to the maximum in total intensity.

Jet Formation. Within the framework of General Relativity, GRMHD 2D axisymmetric simulations of early phases of jet formation from magnetized accretion disks around a non–rotating (Schwarzschild) BH [76] or rotating (Kerr) BH [77] have been performed, assuming a fixed background metric. The magnetic stress induced by the axial magnetic field anchored to the disk (in Keplerian rotation) is responsible of the angular momentum loss of the disk material, triggering the accretion process. In the non–rotating Schwarschild BH case [76] a jet is formed with a maximum velocity of $0.93c$ ($W = 2.7$) if a hydrostatic corona is considered and with a maximum velocity of $0.4c$ if a free falling corona embedding the black hole is assumed. A two layered jet structure (in agreement with theoretical predictions – *e.g.,* [18] –) is found. The inner part is pressure driven and moves at relativistic speeds in the hydrostatic corona case. The outermost part is magnetically driven and subrelativistic. This shell structure might be the origin of the shear layer mentioned in the previous section. Nevertheless, for a fast rotating BH [77], the maximum velocities obtained are $0.4c$ (counter–rotating disk) and $0.3c$ (co–rotating disk), and again the two layered outflow structure is formed (see Fig. 6). As the simulations had to be stopped due to numerical problems quite early [77], the total evolution time was not large enough to develop highly relativistic jets. Therefore, despite of the promising results of Koide's group, the mechanism of jet formation still remains an open and challenging question (see, *e.g.,* [135]).

5.2 Gamma–Ray Bursts

GRBs are known observationally since over 30 years [71]. They consist of very short, non–repeating events (except for a few soft gamma–ray repeaters), with a typical duration between several milliseconds and several hundreds of seconds, showing a large variability even at millisecond scale. They show a bimodal time–distribution, the border between the two groups being at ~ 2 s. The first group is composed of *short* bursts centered around 0.1 s, while the second group consists of *long* bursts (more numerous and softer than the first group) centered at about 15 s. The time–structure is very different from burst to burst.

GRB spectra are non–thermal. The observed energy flux as a function of the energy can be well described by one or a combination of several power laws. The maximum of the energy distribution corresponds to an energy (the energy peak), which is characteristic of each GRB and usually is about several hundreds of keV. The observed fluence on earth is $10^{-5} - 10^{-7}$ erg/cm^2.

For years it was unclear whether GRBs take place at local or cosmological distances (see *e.g.,* [105]). However, a galactic origin can ve excluded, because the BATSE catalog shows an isotropic distribution of GRBs over the sky [104]. BeppoSAX spacecraft [29] has provided accurate coordinates ($\sim$ arc minutes) of the fading X–ray counterparts of GRBs, which has allowed for subsequent ground based observations of faint GRB afterglows at optical and radio wavelengths. The recent redshift determinations, obtained from the optical spectra, prove the cosmological origin at least of the majority of GRBs (see [48] for more information). Observed redshift values are in the range $0.7 \leq z \leq 3.4$ implying

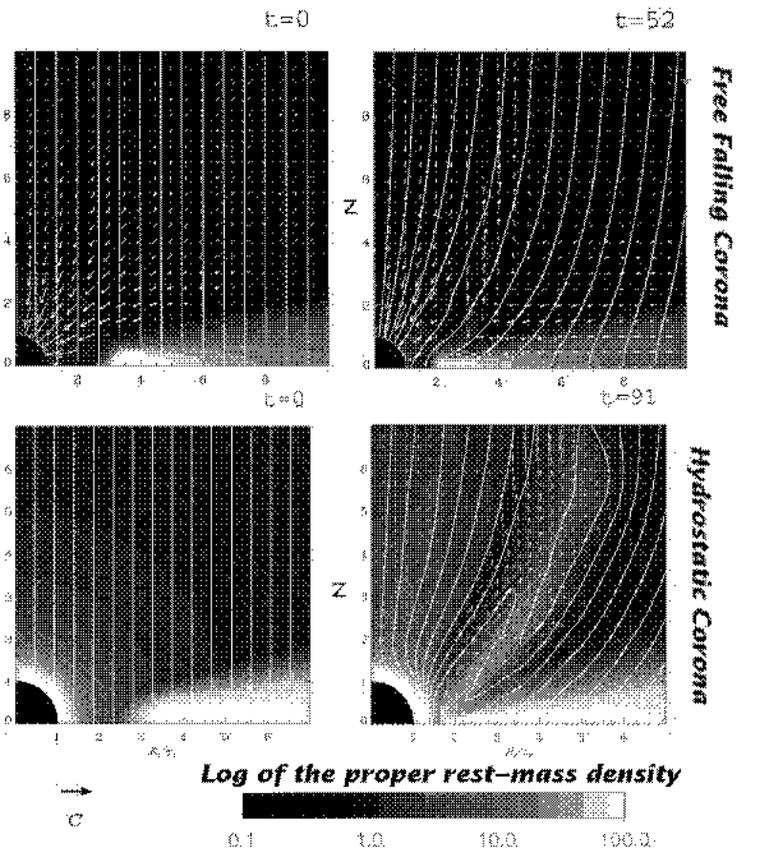

Fig. 6. Figure from [76]. Initial and final time snapshots of the logarithm of the density around a non rotating Schwarzschild BH. Top panels: free falling corona case. Bottom panels: hydrostatic corona case. The solid lines are the magnetic field lines. The vector plots show the flow velocity. On the initial state a uniform axial magnetic field is set up. The disk (in white color) rotates around the BH with Keplerian velocity. In the right panels the jet is formed almost along the magnetic field lines.

emitted gamma-ray energies of $2 \times 10^{51} \leq E \leq 2.3 \times 10^{54}$ erg for an isotropically radiating source. The cosmological origin of the GRBs is consistent with the distribution of bursts in the $\log N - \log P$ plane.

Nonetheless, the accuracy of the positioning is neither sufficient to determine the host galaxies of GRBs nor their progenitors. This picture was challenged by the detection of the Type Ib/c supernova SN 1998bw [46,47] within the error box of GRB 980425 [149,126] whose explosion time and location is consistent with that of the GRB. This suggests a relationship between GRBs and SNe Ib/c, *i.e.*, core collapse supernovae of massive stellar progenitors which have lost their hydrogen and helium envelopes [47,66]. However, the observation of a second fad-

ing X–ray source within the error box of GRB 980425 (different from SN 1998bw) still causes some doubts on the GRB–supernova connection, although the probability of chance coincidence of GRB 980425 and SN 1998bw is almost negligible [126].

Another clue on the nature of the progenitors of GRBs comes from the duration of the shorter bursts and the temporal substructure of the longer bursts ($\sim 1\,\mathrm{msec}$). If the time variation is intrinsic, the length scales involved in the production of a GRB are of about 1 light–millisecond, which in turn points towards compact objects, like NSs or BHs. Furthermore, the non recurrence of the events points towards cataclysmic astrophysical events.

The compact nature of GRB sources, the observed flux and the cosmological distance taken together imply a large photon density and, therefore, a large optical depth for pair production. This is, however, inconsistent with the optically thin source indicated by the non–thermal gamma–ray spectrum, which extends well beyond the pair production threshold at 0.5 MeV. This problem (*compactness problem*) can be resolved by assuming an ultra–relativistic expansion of the emitting region. The bulk Lorentz factor required are $\Gamma > 100$ (see, *e.g.,* [128])

Various catastrophic collapse events have been proposed in order to explain the energies released in a GRB. Among those proposals we find neutron–star/neutron–star mergers [120,52], neutron–star/black–hole mergers [108], collapsars [166,90] and hypernovae [121]. These models rely on the existence of a stellar mass BH hole which accretes several solar masses of matter from a disk (formed during a merger or by a non–spherical collapse) at a rate of $\sim 1\,M_\odot\,\mathrm{s}^{-1}$ [131]. A fraction of the gravitational binding energy released by accretion is converted into neutrino and anti–neutrino pairs, which in turn annihilate into electron–positron pairs. This creates a pair fireball, which will also include baryons present in the environment surrounding the black hole. If the baryon load (the ratio of the fireball mass to its energy) of the fireball is small enough, the baryons are accelerated together with the $\mathrm{e}^+\,\mathrm{e}^-$ pairs to ultra–relativistic speeds with Lorentz factors $> 10^2$ [23,129]. The existence of such relativistic flows is supported by radio observations of GRB 980425 [82]. The rapid temporal decay of several GRB afterglows is inconsistent with spherical (isotropic) blast wave models propagating through the interstellar medium, and instead is more consistent with the evolution of a relativistic jet after it slows down and spreads laterally [137]. Finally, the bulk kinetic energy of the fireball is thought to be converted into gamma–rays via cyclotron radiation and/or inverse Compton processes (see, *e.g.,* [105,128]).

One–dimensional numerical simulations of spherically symmetric relativistic fireballs have been performed by several authors to model GRB sources (*e.g.,* [129,123,124,59]). Multi–dimensional modeling of ultra–relativistic jets in the context of GRBs has for the first time been attempted by Aloy *et al.* [7]. Using a collapsar progenitor model (from [90]) they have simulated the propagation of an axisymmetric jet through the mantle and envelope of a collapsing massive star using a version of GENESIS [3] that includes a background Schwarzschild metric. The jet forms as a consequence of an assumed energy deposition rate

of $10^{50} - 10^{51}$ erg/sec within a 30° cone around the rotation axis. When the jet reaches the surface of the stellar progenitor, the maximum Lorentz factor attained by the flow is about 20. The latter fact implies that Newtonian simulations of this phenomenon [90] are clearly inadequate. The simulations also try to address the ulterior acceleration of the fireball when the jet propagates through an atmosphere of declining density. At the end of the simulations (when the jet has gone over $\sim 10^{11}$ cm) the maximum Lorentz factor is about 50 (for an energy deposition rate of 10^{51} erg/sec). The baryonic contamination is very heterogenous having an average value of 1. However, there are regions (coincident with the parts of the flow having the largest Lorentz factor) where this value is as small as 10^{-5}, which is in agreement with the theoretical expectations (see above). Although the final Lorentz factor is small compared with the predictions of the standard model [23,129], the distance up to which the jet propagation has been tracked (8×10^{10} cm) is much smaller than the one assumed to produce efficient internal shocks ($10^{12} - 10^{14}$ cm) and the fireball to become optically thin ($\sim 10^{13}$ cm, [128]). Therefore, there is still room for further acceleration of the jet until it becomes transparent. Finally, there are recent claims pointing to the possibility of GRB generation without extremely high Lorentz factors and with much smaller masses than in the standard model [153].

6 Summary

Hydrodynamic relativistic processes are on the basis of a variety of challenging astrophysical phenomena. On the other hand, relativistic astrophysics has benefited of the recent development of accurate numerical techniques which have allowed, for the first time, the simulation of ultrarelativistic multidimensional flows. Two main applications are currently addressed. One is in the field of relativistic jets, where very important advances have been made in problems like the jet formation mechanisms or the nature of superluminal sources. The other main application is in the field of GRBs. In order to stress the importance of relativistic hydrodynamic simulations in these fields, let us remind that the generation of both relativistic jets and GRBs is hidden to present observations making numerical simulations the only means to confront the theoretical models.

In the present review we have rewritten the equations governing the dynamics of relativistic perfect fluids paying special attention to their conservative and hyperbolic character (two properties which are extensively exploited by modern numerical techniques). We have also reviewed the evolution of numerical relativistic hydrodynamics, that started as a branch of relativistic astrophysics more than thirty years ago with the pioneering simulations of May and White of stellar core collapse, and which has culminated in the last decade with the introduction of high–resolution shock–capturing (HRSC) methods. The basics of HRSC methods have been summarized, too.

Most of the problems that may be treated by means of numerical simulations require a multidimensional modeling of the flows. Fully three–dimensional simulations in relativistic hydrodynamics are particularly challenging because of

their intrinsic, technical difficulties. These difficulties have been addressed only by a few scientific groups, so far. We have listed those codes that have been used in 3D special or general RHD simulations. Two of them (Cactus and Shibata–Nakamura codes) have implemented a full consistent evolution of the metric of the space–time coupled to the GRHD equations. This fact has allowed them to study, for the first time, problems as complex as the coalescence of simplified models of NSs. The other two codes (GENESIS and Koide codes) allow for a static background metric of the space–time and even for the inclusion of the magnetic field in the equations (Koide's code).

The last Section has been devoted to outline the main results in the simulation of relativistic extragalactic jets and GRBs.

Acknowledgements

This work has been supported in part by the Spanish DGES (grant PB 97-1432). M.A. Aloy thanks to MPA for financial support under its guest program and to the Spanish MEC for the a grant (EX 00 22566499). The authors thank J.M. Ibáñez and E. Müller for the critical reading of the manuscript.

References

1. C.E. Akujor: Astron. Astrophys. **259**, L61 (1992)
2. M. Alcubierre, B. Bruegmann, *et al.* :Phis. Review D **62**, 044034 (2000)
3. M.A. Aloy, J.M$^{\underline{a}}$. Ibáñez, J.M$^{\underline{a}}$. Martí, E. Müller: Astrophys. J. Suppl. **122**, 151 (1999)
4. M.A. Aloy, J.M$^{\underline{a}}$. Ibáñez, J.M$^{\underline{a}}$. Martí, J.L. Gómez, E. Müller: Astrophys. J. Lett. **523**, L125 (1999)
5. M.A. Aloy, J. Pons, J.M$^{\underline{a}}$. Ibáñez, Comput. Phys. Commun. **120**, 115 (1999)
6. M.A. Aloy, J.M$^{\underline{a}}$. Ibáñez, J.M$^{\underline{a}}$. Martí, J.L. Gómez, E. Müller: Astrophys. J. Lett. **528**, L85 (2000)
7. M.A. Aloy, E. Müller, J.M$^{\underline{a}}$. Ibáñez, J.M$^{\underline{a}}$. Martí, A. MacFadyen: Astrophys. J. Lett. **531**, L119 (2000)
8. A.M. Anile. *Relativistic Fluids and Magnetofluids*, (Cambridge University Press, Cambridge 1989)
9. R. Arnowitt, S. Deser, C.W. Misner. *Gravitation: An Introduction to Current Research*, ed., Witten, L., (John Wiley, New York, 1962) p. 227
10. J.M. Attridge, D.H. Roberts, J.F.C. Wardle: Astrophys. J. Lett. **518**, L87 (1999)
11. D.S. Balsara: J. Comp. Phys. **114**, 284 (1994)
12. F. Banyuls, J.A. Font, J.M$^{\underline{a}}$. Ibáñez, J.M$^{\underline{a}}$. Martí, J.A. Miralles: Astrophys. J. **476**, 221 (1997)
13. M.C. Begelman, R.D. Blandford, M.J. Rees: *Rev. Mod. Phys.*, **56**, 255 (1984)
14. M.C. Begelman, M.J. Rees, M. Sikora: Astrophys. J. Lett. **429**, L57 (1994)
15. M.J. Berger, P. Colella: J. Comp. Phys. **82**, 64 (1989)
16. G.V. Bicknell. In *Energy Transport in Radio Galaxies and Quasars*, eds. P.E. Hardee, A.H. Bridle, J.A. Zensus. (ASP Conference Series, 1996) **100**, p. 253
17. R.D. Blandford. In *Active Galactic Nuclei*, eds. T.K.-L. Courvoisier and M. Mayor (Springer, Berlin, 1990), p. 161

18. R.D. Blandford, D.G. Payne: Mon. Not. Roy. Astronom. Soc. **199**, 883 (1982)
19. R.D. Blandford, M.J. Rees: Mon. Not. Roy. Astronom. Soc. **169**, 395 (1974)
20. J.P. Boris, D.L. Book: J. Comp. Phys. **11**, 38 (1973)
21. A.H. Bridle, D.H. Hough, *et al.*: Astronom. J. **108**, 766 (1994)
22. More information can be found under the URL: `http://www.cactuscode.org`.
23. G. Cavallo, M.J. Rees: Mon. Not. Roy. Astronom. Soc. **183**, 359 (1978)
24. J. Centrella, J.R. Wilson: Astrophys. J. **54**, 229 (1984)
25. A.J. Chorin: J. Comp. Phys. **22**, 517 (1976)
26. E. Chow, J.J. Monaghan: J. Comp. Phys. **134**, 296 (1997)
27. P. Colella: SIAM J. Sci. Comput. **3**, 76 (1982)
28. P. Colella, P.R. Woodward: J. Comp. Phys. **54**, 174 (1984)
29. E. Costa, *et al.*: Nature **387**, 783 (1997)
30. R. Courant, K.O. Friedrichs. *Supersonic Flows and Shock Waves*, (Springer, Berlin 1976)
31. W. Dai, P.R. Woodward: SIAM J. Sci. Comput. **18**, 982 (1997)
32. S.F. Davis: *ICASE Report No. 84–20* (1984)
33. G.C. Duncan, P.A. Hughes: Astrophys. J. Lett. **436**, L119 (1994)
34. G.C. Duncan, P.A. Hughes, J. Opperman. In *Energy Transport in Radio Galaxies and Quasars*, eds. P.E. Hardee, A.H. Bridle and J.A. Zensus, ASP Conference Series (1996), **100**, p. 143-148
35. R. Donat, J.M. Font, J.Mª. Ibáñez, A. Marquina: J. Comp. Phys. **146**, 58 (1998)
36. R. Donat, A. Marquina: J. Comp. Phys. **125**, 42 (1996)
37. B. Einfeldt: SIAM J. Numer. Anal. **25**, 294 (1988)
38. F. Eulderink, G. Mellema: Astron. Astrophys. **284**, 654 (1994)
39. F. Eulderink, G. Mellema: Astron. Astrophys. Suppl. **110**, 587 (1995)
40. C.R. Evans. In *Dynamical Space–Times and Numerical Relativity*, ed. J. Centrella, (Cambridge University Press, Cambridge, 1986) p. 3
41. S.A.E.G. Falle, S.S. Komissarov: Mon. Not. Roy. Astronom. Soc. **278**, 586 (1996)
42. B.L. Faranoff, J.M. Riley: Mon. Not. Roy. Astronom. Soc. **167**, 31 (1974)
43. J.A. Font, J.Mª. Ibáñez, J.Mª. Martí, A. Marquina: Astron. Astrophys. **282**, 304 (1994)
44. J.A. Font, M. Miller, W.-M. Suen, M. Tobias: *Phys. Rev.* **D** 61, 044011 (1999)
45. J.A. Font, *Living Reviews in Relativity* (2000), under the URL: `http://www.livingreviews.org/Articles/Volume3/2000-2font`
46. T.J. Galama, P.M. Vreeswijk, E. Pian, F. Frontera, V. Doublier, J.-F. Gonzalez: *IAU Circ.*, 6985 (1998)
47. T.J. Galama, P.M. Vreeswijk, *et al.*: Nature **395**, 670 (1998)
48. T.J. Gallama, (2000), this volume.
49. G. Ghisellini, P. Padovani, A. Celotti, L. Maraschi: Astrophys. J. **407**, 65 (1993)
50. J. Glimm: Comm. Pure Appl. Math. **18**, 697 (1965)
51. S.K. Godunov: *Mat. Sb.*, **47**, 271 (1959)
52. J. Goodman: Astrophys. J. Lett. **308**, L47 (1986)
53. J.L. Gómez, A. Alberdi, J.M. Marcaide: Astron. Astrophys. **274**, 55 (1993)
54. J.L. Gómez, A. Alberdi, J.M. Marcaide: Astron. Astrophys. **284**, 51 (1994)
55. J.L. Gómez, J.Mª. Martí, A.P. Marscher, J.Mª. Ibáñez, J.M. Marcaide: Astrophys. J. Lett. **449**, L19 (1995)
56. J.L. Gómez, J.Mª. Martí, A.P. Marscher, J.Mª. Ibáñez, A. Alberdi: Astrophys. J. Lett. **482**, L33 (1997)
57. J.L. Gómez, A.P. Marscher, A. Alberdi, D.C. Gabuzda: Astrophys. J. **519**, 642 (1999)

58. J.L. Gómez, (2000), this volume.
59. J. Granot, M. Miller, T. Piran, W.-M. Suen. In *Gamma–Ray Bursts*. Eds. R.M. Kippen, R.S. Mallozzi and G.J. Fishman, (AIP conference proceedings 2000), **526**, 540.
60. A. Harten: *Comp. Pure Appl. Math.* **48**, 12 (1995)
61. A. Harten: SIAM J. Numer. Anal. **21**, 1 (1984)
62. A. Harten, B. Engquist, S. Osher, S. Chakravarthy: J. Comp. Phys. **71**, 231 (1987)
63. J.F. Hawley, L.L. Smarr, J.R. Wilson: Astrophys. J. Suppl. **55**, 211 (1984)
64. J.R.A. Hooimeyer, P.D. Barthel, R.T. Schilizzi, G.K. Miley: Astron. Astrophys. **261**, 25 (1992)
65. C.A. Hummel, *et al.*: Astron. Astrophys. **257**, 489 (1992)
66. T.J. Iwamoto, P.A. Mazzali, *et al.*: Nature **395**, 672 (1998)
67. T.W. Jones. In *Supermassive black holes*, (Cambridge University Press, Cambridge 1988), p. 59
68. T.W. Jones, S.L. O'Dell: Astrophys. J. **214**, 522 (1977)
69. W. Junor, J.A. Bireta: Astronom. J. **109**, 500 (1995)
70. A. Kheyfets, W.A. Miller, W.H. Zurek: *Phys. Rev. D* **41**, 451 (1990)
71. R.W. Klebesadal, I.B. Strong, R.A. Olsen: Astrophys. J. Lett. **182**, L85 (1973)
72. A. Königl: Astrophys. J. **243**, 700 (1981)
73. S. Koide: Astrophys. J. **487**, 66 (1997)
74. S. Koide, K.-I. Nishikawa, R.L. Muttel: Astrophys. J. Lett. **463**, L71 (1996)
75. S. Koide, K. Shibata, T. Kudoh: Astrophys. J. Lett. **495**, L63 (1998)
76. S. Koide, K. Shibata, T. Kudoh: Astrophys. J. **522**, 727 (1999)
77. S. Koide, D.L. Meier, K. Shibata, T. Kudoh: Astrophys. J. **536**, 668 (2000)
78. S.S. Komissarov: *Sov. Astron. Lett.* **16**(4), 284 (1990)
79. S.S. Komissarov, S.A.E.G. Falle: Mon. Not. Roy. Astronom. Soc. **288**, 833 (1997)
80. S.S. Komissarov, S.A.E.G. Falle: Mon. Not. Roy. Astronom. Soc. **297**, 1087 (1998)
81. A.K. Kembhavi, J.V. Narlikar. *Quasars and Active Galactic Nuclei. An introduction* (Cambridge University Press, Cambridge 1999)
82. S.R. Kulkarni, D.A. Frail, *et al.*: Nature **395**, 663 (1998)
83. P. Laguna, W.A. Miller, W.H. Zurek: Astrophys. J. **404**, 678 (1993)
84. N.K. Lahy, N.K. *A Particle Method for Relativistic Fluid Dynamics*, M.Sc. Thesis, Monash University (1989)
85. R.A. Laing. In *Radio Galaxies and Quasars*, eds. P.E. Hardee, A.H. Bridle, J.A. Zensus. ASP Conference Series, Vol. 100 (1996) p. 241
86. W. Landry, S.A. Teukolsky: submitted (1999). See also: gr-qc/9912004
87. R.J. LeVeque. *Numerical Methods for Conservation Laws*, 2nd. edn. (Birkhäuser, Basel 1992)
88. A.P. Lovanov, T.P. Krichbaum, *et al.*: Astron. Astrophys. **340**, L60 (1998)
89. L.B. Lucy: Astronom. J. **82**, 1013 (1977)
90. A. MacFadyen, S.E. Woosley: Astrophys. J. **524**, 262 (1999)
91. P.J. Mann: Comput. Phys. Commun. **67**, 245 (1991)
92. P.J. Mann: Comput. Phys. Commun. **107**, 188 (1993)
93. A. Marquina: SIAM J. Sci. Comput. **15**, 892 (1994)
94. A.P. Marscher: Astrophys. J. **235**, 386 (1980)
95. A.P. Marscher, W.K. Gear: Astrophys. J. **298**, 114 (1985)
96. A.P. Marscher, W.K. Gear, J.P. Travis. In *Variability of Blazars*, ed. by E. Vataoja and M. Valtonen (Cambridge University Press, Cambridge 1992)

97. J.M^{a}. Martí, J.M^{a}. Ibáñez, J.A. Miralles: Phis. Review D **43**, 3794 (1991)
98. J.M^{a}. Martí, E. Müller: J. Fluid Mech. **258**, 317 (1994)
99. J.M^{a}. Martí, E. Müller, J.M^{a}. Ibáñez: Astron. Astrophys. **281**, L9 (1994)
100. Martí, J.M^{a}, E. Müller, J.A. Font, J.M^{a}. Ibáñez: Astrophys. J. **448**, L105 (1995)
101. Martí, J.M^{a}, Müller, E., J.A. Font, J.M^{a}. Ibáñez, A. Marquina: Astrophys. J. **479**, 151 (1997)
102. J.M^{a}, Martí, E. Müller: *Living reviews in Relativity* (1999), under the URL: `http://www.livingreviews.org/Articles/Volume2/1999-3marti`
103. M.M. May, R.H. White: *Methods Comput. Phys.* **7**, 219 (1967)
104. C.A. Meegan, G.J. Fishman, *et al.*: Nature **355**, 143 (1992)
105. P. Mèszáros. In *Proc. of the 17th Texas Symp. on Relativistic Astrophysics and Cosmology*, eds. H. Böhringer, G.E. Morfill and J.E. Trümper, (N. Y. Acad. Sci., 1995) p. 440
106. A.J. Mioduszewski, P.A. Hughes, G.C. Duncan: Astrophys. J. **476**, 649 (1997)
107. I.F. Mirabel, L.F. Rodriguez: Nature **371**, 46 (1994)
108. R. Mochkovitch, M. Hernanz, J. Isern, X. Martin: Nature **361**, 236 (1993)
109. J.J. Monaghan: Comput. Phys. Rep. **3**, 71 (1985)
110. J.J. Monaghan: J. Comp. Phys. **136**, 298 (1997)
111. T. Nakamura: Prog. Theor. Phys. **65**, 1876 (1981)
112. T. Nakamura, K. Oohara, Y. Kojima: Prog. Theor. Phys. Suppl. **90**, 76 (1987)
113. T. Nakamura, K. Oohara. In *Numerical Astrophysics*. Ed. by S.M. Miyama, K. Tomisaka and T. Hanawa (Kluwer Academic, Astrophysics and space science library 1999) **240**, p.247
114. T. Nakamura, K. Maeda, S. Miyama, M. Sasaki: Prog. Theor. Phys. **63**, 1229 (1980)
115. K.-I. Nishikawa, S. Koide, *et al.*: Astrophys. J. **483**, L45 (1997)
116. K.-I. Nishikawa, S. Koide, *et al.*: Astrophys. J. **498**, 166 (1998)
117. M.L. Norman, L. Smarr, K.-H.A. Winkler, M.S. Smith: Astron. Astrophys. **113**, 285 (1982)
118. M.L. Norman, K.-H.A. Winkler. In: *Astrophysical Radiation Hydrodynamics*, ed. by M.L. Norman, K.-H.A. Winkler (Reidel, Dordrecht, 1986) p. 227
119. S. Osher, S. Chakravarthy: SIAM J. Numer. Anal. **21**, 995 (1984)
120. B. Pacyński: Astrophys. J. **308**, L43 (1986)
121. B. Pacyński: Astrophys. J. **494**, L45 (1998)
122. A. Panaitescu, L. Wen, P. Laguna, P. Mészáros: Astrophys. J. **482**, 942 (1999)
123. A. Panaitescu, P. Mészáros: Astrophys. J. **492**, 683 (1998)
124. A. Panaitescu, P. Mészáros: Astrophys. J. **526**, 707 (1999)
125. L.I. Petrich, S.L. Shapiro, R.F. Stark, S.A. Teukolsky: Astrophys. J. **336**, 313 (1989)
126. E. Pian, L. Amati, *et al.*: Astrophys. J. **538**, 638 (2000)
127. T. Piran: J. Comp. Phys. **35**, 254 (1980)
128. T. Piran: *Physics Reports* **314**, 575 (1999)
129. T. Piran, A. Shemi, R. Narayan: Mon. Not. Roy. Astronom. Soc. **263**, 861 (1993)
130. J.A. Pons, J.M^{a} Martí, E. Müller: J. Comp. Phys. **422**, 125 (2000)
131. R. Popham, S.E. Woosley, C. Fryer: Astrophys. J. **518**, 356 (1999)
132. P.L. Roe: J. Comp. Phys. **43**, 357 (1981)
133. R.D. Richtmyer, K.W. Morton: *Difference Methods for Initial–value Problems*, (Wiley–Interscience, New York 1967)
134. Rosen, A., Hughes, *et al.*: Astrophys. J. **516**, 729 (1999)
135. C. Sauty (2000), this volume.

136. R.H. Sanders, K.H. Prendergast: Astrophys. J. **188**, 489 (1974)
137. R. Sari, T. Piran, J.P. Halpern: Astrophys. J. Lett. **519**, L17 (1999)
138. P.A.G. Scheuer: Mon. Not. Roy. Astronom. Soc. **166**, 513 (1974)
139. V. Schneider, U. Katscher, *et al.*: J. Comp. Phys. **105**, 92 (1993)
140. M. Shibata, T. Nakamura: Phis. Review D **52**, 5428 (1995)
141. M. Shibata: Prog. Theor. Phys. **101**, 251 (1999)
142. M. Shibata: Prog. Theor. Phys. **101**, 1199 (1999)
143. M. Shibata: Phis. Review D **60**, 104052 (1999)
144. M. Shibata, T.W. Baumgarte, S.L. Shapiro: Phis. Review D **61**, 044012 (2000)
145. S. Siegler, H. Riffert: Astrophys. J. Suppl. **531**, 1053 (2000)
146. C.W. Shu:*Math. Comp.* **49**, 105 (1987)
147. C.W. Shu, S.J. Osher: J. Comp. Phys. **83**, 32 (1989)
148. L. Smarr, J.R. Wilson, R.T. Barton, R.L. Bowers: Astrophys. J. **246**, 515 (1981)
149. P. Soffitta, M. Feroci, *et al.*: *IAU Circ.* 6884 (1998)
150. J. Sollfrank, P. Huovinen, M. Kataja, P.V. Ruuskanen, M. Prakash, R. Venugopalan: *Phys. Rev. C* **55**, 392 (1997)
151. R.F. Stark, T. Piran: Comput. Phys. Rep. **5**, 221 (1987)
152. M.R. Swain, A.H. Bridle, S.A. Baum: Astrophys. J. **507**, L29 (1998)
153. J.C. Tan, C.D. Matzner, C.F. McKee: Astrophys. J. , submitted, (2000)
154. M. Taub: *Phys. Rev.* **74**, 328 (1948)
155. K.W. Thomson: J. Fluid Mech. **171**, 365 (1986)
156. S.J. Tingay, D.L. Jauncey, R.A. Preston, J.E. Reynolds, D.L. Meier *et al.*: Nature **374**, 141 (1995)
157. E. Toro: *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*, 1st edn. (Springer, Berlin 1997)
158. B. van Leer: J. Comp. Phys. **32**, 101 (1979)
159. M.H.P.M. van Putten: J. Comp. Phys. **105**, 339 (1993a)
160. M.H.P.M. van Putten: Astrophys. J. Lett. **408**, L21 (1993b)
161. M.H.P.M. van Putten: Astrophys. J. Lett. **467**, L57 (1996)
162. J. von Neumann, R.D. Richtmyer: *J. Appl. Phys.* **21**, 232 (1950)
163. L. Wen, A. Panaitescu, P. Laguna: Astrophys. J. **486**, 919 (1997)
164. J.R. Wilson: Astrophys. J. **173**, 431 (1972)
165. J.R. Wilson. In: *Sources of Gravitational Radiation*, ed. by L.L. Smarr, (Cambridge University Press, Cambridge 1979) p. 423
166. S.E. Woosley: Astrophys. J. **405**, 273 (1993)
167. Yang, J.Y., Chen, M.H., Tsai, I-N. and J.W. Chang: J. Comp. Phys. **136**, 19 (1997)

Epilogue

Roger Blandford

130-33 Caltech, Pasadena, CA 91125

Abstract. There are now several types of relativistic flows in astrophysical settings. The foremost examples are jets and disks orbiting spinning black holes, pulsar winds and gamma ray bursts. As discussed at this meeting, these flows exhibit unusual kinematic and dynamical properties, that distinguish them from non-relativistic flows. It is possible that all collimated outflows are essentially hydromagnetic or electromagnetic. Future study of relativistic flows will rely heavily on numerical experiments. Model relativistic flows provide a basis for carrying out secondary studies of the underlying plasma physics, particle acceleration, magnetic field amplification and the emission and transfer of radiation, particularly at shock fronts. Some current opportunities in observation, phenomenology and theory are briefly suggested.

1 Relativistic Flows in Astrophysics

Although the initial forays into relativistic gas dynamics [12] and relativistic MHD [4] were mainly stimulated by applied mathematical curiosity, astronomical observations now provide abundant incentive to consider relativistic flows in detail. In this brief summary, I shall review these developments in rough, historical order, emphasise some contemporary problems and suggest some future directions. I refer to the many excellent presentations at this workshop for detailed discussions and more extensive bibliographies.

1.1 AGN

Ever since the pioneering, VLBI observations in the early 1970's [1], [13] we have known that compact, extragalactic radio sources can expand with space velocities within one percent of the speed of light so as to create a strong "superluminal" illusion [9]. The emitting features, in variable, compact radio sources, were soon associated with external and internal shock waves and the jet-like, as opposed, to spherical character of the flows was established. Observations of jets at optical, X-ray and, especially γ-ray energies followed and much effort is currently being applied to understanding the details of how different types of jet emit throughout the electromagnetic spectrum. The general picture that has emerged [Agudo, Celotti, Close, Georganopoulos, Kovalev, Mastichiadis, Malzac] [1] is that relativistic electrons are accelerated, mainly at shock fronts, and that

[1] See *Similarities and Universality in Relativistic Flows*, Proceedings of the PhD Euroconference, Logos Verlag, Eds. M. Georganopoulos et al., 2001, for a written form of the talks given at SURF 2000 cited here – *the eds.*

Axel W. Guthmann et al. (Eds.): Relativistic Flows in Astrophysics, LNP 589, pp. 227–237, 2002.

they emit synchrotron radiation at low frequency and inverse Compton emission at high frequency, with the former providing the soft photons for the latter in low power sources and the disk ultraviolet emission supplying the soft photons in high power sources.

Although early, relativistic jet models were essentially gas dynamical, with the magnetic field evolving passively, it is now widely believed that jets are collimated by large-scale magnetic field [Casse, Ferreira, Nitta] with a disk angular velocity vector determining the jet axis. However the details are controversial. Most models of jet formation and collimation that have been published to date are non-relativistic, analytic MHD solutions where the intrinsic anisotropy of the Maxwell stress tensor is ultimately responsible for the collimation. This is reasonable because, as it is generally reasoned, what is being described is a collimating, magnetic sleeve that confines an ultrarelativistic flow that emanates from the black hole and the inner disk. However, some authors have argued that the magnetic field is always primarily poloidal and any toroidal field that is generated by the rotation will quickly vanish through reconnection. At the other extreme it has been proposed that rotation dominates and the magnetic field lines behave like a coiled spring pushing the jets out along the spin axis. Both components are relevant to centrifugal models where the inertia of the outflowing plasma plays a crucial role. Finally, there are models where there is no long range order to the field and it is the local anisotropy associated with loops of magnetic flux that create the collimation. The stability of most of these models, particularly to non-axisymmetric modes, is only recently coming under scrutiny [Lery].

Yet another debate, and this is probably the central phenomenological question in interpreting relativistic flows, is "What is the working substance?". [Markowith, Yamasaki]. There is far too large a radiation density around the inner disk for the flows to start life as just pairs and there has to be some other carrier of momentum. One possibility, considered in some early models was the radiation itself. However, elementary kinematic considerations make it quite unlikely that large Lorentz factors can be achieved. Protons, whose radiative efficiency is very low, could be responsible, though it is very hard to see how they could be accelerated efficiently in a beam. In my view, the most reasaonable candidate is electromagnetic field. Note, that I am only suggesting that jets start off in a predominately electromagnetic form. A quite likely sequence is that, at some finite distance from the black hole where annihilation cannot keep up with production, jets metamorphose into a pair plasma. This may be where the observed γ-rays are produced. At a yet greater distance, these jets should ultimately interact strongly with their surroundings as they become radio sources and decelerate. Presumably, when the jet is powerful, the outflow can remain relativistic and we have a FR2 source; when the jet is weak and decelerates to a subsonic speed, a FR1 source is formed. At this point, we are dealing with reasonably well-resolved extended radio sources and now have a much better understanding of the physical conditions in the surrounding gas. This should enable us to make more confident descriptions of radio source evolution and more

quantitative estimates of the total jet powers [Blundell, Manolakou, Polatidis], both in individual sources (where they can exceed the bolometric power of the AGN) and collectively as a contribution to the luminosity density of the universe.

Although jets are certainly the most dramatic, relativistic flows associated with AGN, gas in the accretion disk also moves with mildly relativistic speed and general as well as special relativity must be invoked to account for the X-ray line spectroscopy - a quite dramatic vindication of the black hole model. This is possible evidence that the holes are spinning, because the prominent red wings that are sometimes observed should only be formed if the disk can approach the horizon and this only happens for a geometrically thin disks orbiting a rapidly spinning hole in a prograde sense [11].

1.2 Pulsar Winds

Pulsars were discovered soon after there was evidence for relativistic effects in AGN and it was quickly realized that they should also be (relativistic) electromagnetic objects [7][2]. Originally, the field was thought to be that of a rotating, vacuum, magnetic dipole, though it was soon realized that the magnetosphere had to contain plasma which would at least seriously modify the electromagnetic field and might have dynamical importance [3]. A similar metamorphosis of energy from mechanical, through electromagnetic, pairs and ions is envisaged though the details of how and where these transformations occur are no less controversial than they are with AGN jets despite much valiant, theoretical effort.

Important Chandra observations of the Crab Nebula, and a few other plerions, have demonstrated that the presumed spherical winds actually exhibit "jets", giving the lie to the assertion that disks are necessary for jet formation. (Actually, they also appear to possess features that look like "disks", though these are probably equatorial current sheets, like those found to be associated with the solar wind.)

1.3 Galactic Superluminal Sources

The association of black hole accretion disks with AGN led to the (morphological) expectation that binary X-ray sources (where the direct evidence for disks was stronger than in AGN) should also produce jets. The early evidence (eg in Sco-X1) was confused, but with the discovery of the jets in SS433 [5] [Rowell], (where the jet velocity and its variation could be accurately measured) the matter was settled. However, here, and in all other known examples to date [Fender], the outflow speed is only mildly relativistic in contrast to what has been found with the AGN jets. Indeed, there is no dynamical or kinematical objection to the jets being created by radiation pressure and as these sources are operating quite close to the Eddington limit, radiation cannot be ignored. These inner disks, like those associated with AGN accreting close to the Eddington rate, comprise radiation-dominated gas.

The Galactic superluminal sources [6], which are believed to contain black holes, also exhibit "Quasiperiodic Oscillations", or QPOs, analogous to the oscillations that have been reported from neutron star systems. The modulation is generally associated with standing modes in the relativistic accretion disk. However, the disk can only be the clock; the X-ray emission is so hard that it must actually originate in the corona. This is additional, circumstantial evidence for a strong magnetic coupling between the disk and its surroundings and, in principle, a strong diagnostic of general relativistic disk flow.

1.4 Gamma-Ray Bursts

Although early discussions of the nature of GRBs clearly recognized the implications if they were at cosmological distances [10], it wasn't until the BATSE catalog was produced that it became clear that this was probably the case and that GRBs probably expanded even faster than AGN jets [8]. The inference was verified by the measurement of afterglow redshifts and the discovery of radio scintillation[Downes, Kobayashi, Galama, Sari]. It now seems to be generally accepted that bulk Lorentz factors, variously estimated as lying in the range $100 < \Gamma < 1000$, are required to avoid pair production by the escaping high energy γ-rays. The gamma ray burst itself is most commonly associated with the dissipation of internal shocks that form in the expanding fireball and reflect variation in the source over a relatively long timescale ~ 100 s for the better studied "long" bursts. The afterglow, which can be traced for over year in some cases, is associated with a blast wave, initially ultrarelativistic, formed by the swept up interstellar medium. The evidence that this flow is non-spherical, ie that GRBs are also jets – the observation of achromatic spectral breaks and a desire to limit the explosion energy – is improving but is not yet decisive.

The study of the dynamics and radiative properties of afterglows has partly recapitulated the study of AGN jets, although there is now an impressively detailed phenomenological description of comprehensive observations of over 20 bursts throughout the electromagnetic spectrum. Although there is circumstantial evidence that the long bursts are associated with star-forming regions in galaxies, the nature of the sources is still unclear. We know less about the short bursts, though the soft gamma repeaters are probably associated with magnetars.

In most contemporary models of the non-repeating bursts, a black hole is either formed or augmented. Many of these models, specifically the collapsar models, raise fundamental questions of relativistic gas dynamics, including the question of whether or not we are dealing with a fluid at all! The afterglows raise the same questions that came up with AGN concerning particle acceleration and field amplification. "How are particles accelerated and how is field amplified?". From an astronomical perspective, we also want to understand the place of GRBs in the scheme of advanced stellar evolution and supernova explosions as well as their potential as sources of neutrinos and gravitational radiation as well as their "environmental impact".

1.5 Other Relativistic Flows

There are several other types of mildly relativistic flow that have been considered in astrophysics including accretion onto neutron stars, broad absorption line quasar outflows and, most fundamentally of all, early universe cosmology which is also an exercise in relativistic gas dynamics with a transition from radiation-dominance to gas-dominance, just like that in accretion disks!

2 Relativistic Flows

2.1 Gas Dynamics

One of many satisfying features of both special and general relativity is how harmoniously they accommodate gas dynamics. The relativistic formalism emphasizes symmetry and conservation laws in a manner that is sometimes lost in the more engineering-oriented development of the non-relativistic subject. Gravity can be ignored for application to jets, winds and GRBs and the governing equations, derivable from setting the divergence of the mass particle current vector and the stress-energy tensor to zero, express the conservation of mass, momentum and energy. This leads to counterparts of familiar non-relativistic descriptions for ID flow, shock discontinuities and so forth. In many analyses of relativistic flows, the fluid is often taken to be isotropic and ultrarelativistic, that is to say the pressure is dominated by radiation or high energy leptons with an internal sound (proper) speed of $2^{-1/2}c$. In this case the effective Mach number is $M = 2^{1/2}u$ where u is the proper bulk speed. In other applications, non-relativistic protons are also present and reduce the sound speed.

However, there are some serious worries as to how complete a description this really is in many of the environments where these results are applied. For example, in a pair creation region, mass will not be conserved. Furthermore, momentum and energy will not be conserved in the presence of inverse Compton scattering. Another worry is that shear stress is usually ignored when dealing with jets while it is seen as an intrinsic part of another common astrophysical shear flow, the accretion disk. When we make 1D jet models we are implicitly assuming that jets are enclosed by narrow, turbulent, boundary layers that do not spread so that Mach numbers can attain large values – over 300 in some collapsar models – and the ratio of the bulk kinetic energy to the internal energy exceeds $\sim M^2$. This is supposed to happen naturally with essentially no noise and internal dissipation reconverting the bulk energy to internal energy. An aerodynamicist would think this strange!

High Mach number jets have some unusual properties (both non-relativistically and relativistically). If the fluid starts from a subsonic chamber, where it is all in causal contact, and accelerates through a pair of nozzles to form "twin exhausts" with Mach number much larger than the reciprocal of the jet opening angle θ^{-1}, (as is thought to happen in GRBs), then the different parts of the jet flow will fall out of causal contact. Now, in the case of a GRB, the jet is likely to be preceded by a relativistic blast wave propagating into the

surrounding medium. Initially the different elements of this blast wave will also be out of causal contact. However, as the blast wave decelerates, M will fall to θ^{-1} and transverse causal contact will be re-established. This "hello–goodbye–hello" behavior is thought to be responsible for the achromatic breaks in the afterglow emission and is reminiscent of inflationary cosmology!

Ultrarelativistic flows do have some distinctive kinematic properties which mostly derive from the fact that the 3-speed is limited to that of light. This in turn, leads to strong Doppler-shifting and beaming of the emitted radiation. These effects can be extremely large and can lead to insignificant parts of the source dominating what we observe

2.2 Passive Gravitational Field

The next most complicated class of problems involves gas dynamical flows in the presence of a passive gravitational field. A prime example is an accretion disk in orbit about a black hole. There is now a lot of interest in solving these problems using the full machinery of general relativity. For example, "diskoseismological" oscillation modes have been calculated and their frequencies can be made to match QPO observations.

2.3 Active Gravitational Field

Flows where the spacetime is dynamic are far harder to analyze and numerical methods are necessary. The most pressing examples are neutron star – neutron star/black hole models of GRBs. Simulations have been used to determine the timescales for coalescence and, for example, to show that neutrino emission is unlikely to be very important in driving the burst.

2.4 Magnetohydrodynamics

As I have already remarked, most models of relativistic jets, plerions and disks are intrinsically magnetised. In particular, we now know how magnetic field is amplified in a non-relativistic disk, through the magnetorotational instability and it is now generally agreed that disk evolution is a magnetohydrodynamic problem. Similarly, jet collimation is generally argued to be due to anisotropic magnetic stress on the grounds that the maximum gas pressures allowed by X-ray observational constraints are too small to effect collimation.

In non-relativistic astrophysical MHD [Sauty], it is commonly assumed that the electrical conductivity is infinite, implying that the electric field vanishes in the centre of momentum frame. When this happens the flow evolves under a set of locally deterministic equations which give the partial derivatives of the velocity, density and the magnetic field with respect to time. (There is an applied mathematical nicety involving the degeneracy of the signal speeds along the field. which is probably inconsequential in practice.) Note that there is no need for an equation to describe the temporal evolution of the current density as this is

given by the curl of the magnetic field. The charge density can be determined after the fact from gauss' law, if needed, but it has no dynamical role in the non-relativistic limit.

By contrast, when we try to do the same thing in relativistic MHD, some awkward questions are raised because we cannot ignore the displacement current and the charge density. Maxwell's equations are evolutionary equations for $\rho, \boldsymbol{E}, \boldsymbol{B}$. In the infinite conductivity limit, there is no evolutionary equation for $\boldsymbol{j}$. Only when we introduce a finite conductivity, so that the current density is given by some form of Ohm's law in the center of momentum frame, do we fix the current locally. However, when the conductivity is so large that there is insignificant electric potential difference along the magnetic field, the current flow must be determined by what happens elsewhere, in regions where there is dissipation or by the boundary conditions.

There is a second possible problem with relativistic MHD that can have a large bearing on the outcome. Traditionally there are three modes of wave propagation known as fast and slow magnetosonic propagation, together with the intermediate (or Alfvén) mode. Now the slow mode is determined by the sound speed in the gas. This is traditionally taken to be isotropic and (if the gas has a high temperature) to be $3^{-1/2}c$. However, in many flows, including those around black holes, it is possible for the electrons to cool on a dynamical timescale. The particle distribution function may become highly anisotropic with repect to the ambient magnetic field. In this case the effective sound velocity along the direction of the magnetic field can become arbitrarily close to the speed of light. This, in turn affects the characteristics and has implications for the development of shock waves and the causal structure of relativistic flows.

2.5 Force-Free Electrodynamics

A useful approximation for handling magnetised, relativistic flows, that simplifies the calculation, though does not remove the first of the above difficulties, is to adopt the relativistic force-free approximation, namely that $\rho\boldsymbol{E}+\boldsymbol{j}\times\boldsymbol{B}=0$. This immediately supplies a constitutive relation for the component of the current density resolved perpendicular to the magnetic field, $\rho\boldsymbol{E}\times\boldsymbol{B}/B^2$. The parallel component must be fixed by boundary conditions, just as in the non-relativistic case. In this approximation, which is surely good for field lines which thread a black hole event horizon and quite possibly for pulsar winds, we dispense with the velocity all together. The role of the plasma is to supply charge density and current. The invariant $\boldsymbol{E}\cdot\boldsymbol{B}$ vanishes and there will generally be ample charge to keep the invariant B^2-E^2 positive.

2.6 Radiation-Dominated Gas Flows

Increasing attention is being paid to the dynamics of radiation-dominated fluid. The prime example is the open universe. However, here the vorticity and magnetic field are thought to be quite small. By contrast, relativistic accretion disks and optically thick jets are quite likely to behave very differently. These are, by

definition, shear flows in which magnetic fields grow on a dynamical timescale. However, the field may become quite inhomogeneous in a manner which will facilitate radiative transfer. In fact, photons can transfer momentum as well as energy under some circumstances.

3 Numerical Simulations

It has been made quite apparent here that the way forward is through numerical simulation [Aloy, Peitz]. Impressive advances have been reported in the testing and deployment of large three (and four) dimensional codes with and without magnetic field and improvements in speed and memory make quite sophisticated investigations a practical proposition. The numerical study of radiation-dominated, relativistic magnetohydrodynamics is on the horizon and promises the biggest surprises as we strive to develop some understanding of what really happens to gas accreting onto a black hole. Even in simple flows, breaking spherical or axisymmetry and going beyond self-similarity is producing large changes in our outlook. In addition, the capability to tackle non-linear perturbations - the only sort that observers can see - is crucial.

This, is not to say that there is no further role for analytical approaches. In some sense, they become more important. This is because it is extremely hard to represent the results of multi-dimensional computations, graphically or verbally in a manner that allows one to divine general principles and predict what will happen in other flows. Having a simple description of the most important features of a complex flow is immensely valuable. It is also important to distinguish numerical simulation, which aspires to reproduce an accurate representation of a flow from numerical experiments. As with much experimental physics, a laudable goal of numerical experiments is to get so much insight that it is possible to replace them with a working model that can be used as a subunit of a larger investigation.

Understanding the flow is not the end of the matter. It is important to use the fluid solution to provide a framework to discuss higher order features like the plasma physics, particle acceleration and radiative transfer. This is vital if we want to interpret the diagnostic observations of relativistic sources. Carrying out these secondary studies relativistically is turning out to be no less of a numerical challenge than computing the basic flows.

4 What Now?

4.1 Observation

The observational prospects are good. On AGN jets, there is an opportinity for using polarimetric observations and imaging to tell us if jets have an electromagnetic or a gas dynamical origin and for understanding what factors determine the jet power. Increasingly detailed X-ray observations of sources like M87, Cygnus A and Pictor A are providing excellent laboratories for determining directly,

where relativistic electrons are accelerated, what is the magnetic feld geometry and how the particles are transported. Temporal studies should clarify the extent to which jets can be thought of as continuous flows or a sequence of explosive outbursts. The spectacularly detailed X-ray spectroscopy that is emerging from both XMM-Newton and Chandra will eventually be interpreted and should define the geometry of the gas flow near to black holes and, especially, identify where the energy is dissipated.

Our view of plerions has been considerably enhanced by their association with soft gamma repeaters and magnetars. However there is still a lot of uncertainty in these identifications and an imperfect understanding of how a plerion changes in reponse to a burst. More coordinated observations are needed.

One of the big observational challenges in studying the Galactic superluminal sources is to find a "microblazar" - a high Lorentz factor jet pointed towards us. (It is possible that the recently-discovered, super-Eddington compact sources in nearby galaxies could be of this type.) If, conversely, we can persuade ourselves that these objects do not exist, then it will probably tell us something useful about jet formation.

Turning to GRBs, HETE-2 should identify some short bursts and point to a physical origin for this subclass to complement the tremendous observations of the long bursts that derive from Beppo-SAX and BATSE.

4.2 Phenomenology

The basic jet emission model involving shock fronts admits a simple, testable prediction. The kinematic velocity of the emitting feature will differ from the velocity of the emitting gas, which can be measured through the Doppler shift. It would be very nice to demonstrate this in some sources.

There is great need to discover the true laws of MHD which will describe how collisionless plasmas behave in practice on the largest scales as opposed to the applied mathematical idealisations on which we must rely at present. Undoubtedly our best hope for understanding non-relativistic flows lies with careful analysis of the superb observations of the solar corona by the YOHKOH, SOHO and TRACE spacecrafts as well as the *in situ* observations of the solar wind and planetary magnetospheres. We need to understand how currents flow – are they distributed or filamentary - how shocks create entropy under a wide variety of defining conditions, how much energy is dissipated in reconnection regions and in what form, how turbulent spectra develop and the connection to magnetic amplification by dynamo action. We would like to understand solar flares as a prelude to understanding the energisation of an accretion disk corona and the means of launching the solar wind which is surely relevant to the formation of jets.

The best laboratory that we have for the relativistic flows that are the subject of this meeting is surely the Crab Nebula. Here Chandra and HST observations are changing our view of the pulsar wind/jet, and its termination through a strong shock front. However we still do not have an accepted determination of its

speed and composition. The notion that magnetised flows generically collimate into jets even without a disk, if true, is of immediate relevance to GRBs.

A related question is "How much is the character of the jet dependent upon the central compact object?". We are pretty confident that the Galactic superluminals are identified with black holes but Sco X-1 (and possibly, SS433) may derive from neutron stars. In addition the jets we observe best are formed by protostars, so a "compact object" is far from necessary. This is a good clue as to how relativistic jets are powered. Perhaps all that is required to make a jet large relative angular velocity (in units of the Keplerian value).

Another research frontier is the Ultra High Energy cosmic rays. There is a good chance that, as observations continue to improve over the next few years, we will be forced to a phenomenological model of their origin which will surely involve ultrarelativistic plasma physics.

4.3 Theory

There is now a large backlog of unsolved problems in fundamental theory that must be solved before we can model astrophysical, relativistic flows with confidence. For many of these, as I have emphasized, the requisite computational tools are becoming available. Perhaps the most pressing need, with the largest, general implications, is to understand how magnetic fields evolve around black holes, in the radiation or ion-dominated disks and in the magnetosphere above the hole. We will almost certainly need to perform large scale numerical simulations and will want to use these to determine the relative efficiencies for the release of energy by the hole and the disk and the connection to jets. This problem will be very hard to solve, though it is fairly well posed. In the case of an ion plasma we will want to use plasma simulations to understand better the collisionless coupling between the hot ions and the cool electrons

There are several additional, interesting formal challenges which may turn out to be relevant to interpreting the observations, such as making a theory for the radiative transfer of plasma waves in a curved spacetime, developing a relativistic theory of reconnection and reworking the theory of the MRI for a relativistic disk.

As the diverse contributions to this lively workshop attest, these are only a few of the possibilities inherent in this young and exciting field.

Acknowledgements

I am indebted to Markos Georganopoulos, Axel Guthmann, Konstantina Manolakou and Alexandre Markowith for their invitation to this meeting, their generous hospitality and allowing me to fulfil a lifelong ambition to visit Greece. Support under NASA grant 5-2837 is gratefully acknowledged.

References

1. Cohen, M. H. *et al* 1971 ApJ 170 207
2. Gold, T. Nature 218 731
3. Goldreich, P. & Julian, W. H. 1969 ApJ 157 869
4. Lichnerowicz, A. 1967 Relativistic Hydrodynamics and Magnetohydrodynmics New York:Benjamin
5. Margon, B. *et al* 1979 ApJ 230 L41
6. Mirabel, I. F. & Rodriguez, L. F. 1994 Nature 371 46
7. Pacini, F. Nature 216 567
8. Paczyński, B. 1986 ApJ 308 L51
9. Rees, M. J. 1966 Nature 211 468
10. Ruderman, M. 1975 Ann. N. Y. Acad. Sci. 262 164
11. Tanaka, Y. *et al* 1995 Nature 375 659
12. Taub, A. H. 1948 Phys. Rev. 74 328
13. Whitney, A. R. *et al* 1971 Science 173 225

Index