



ISTITUTO NAZIONALE DI FISICA NUCLEARE
Laboratori Nazionali di Frascati

FRASCATI PHYSICS SERIES

FRASCATI
PHYSICS SERIES

A. Morselli,
P. Picozza



GIULIA
VERNETTO

Consorzio Interuniversitario per la Fisica Spaziale

INTERNATIONAL SCHOOL OF SPACE SCIENCE, 2001 Course on

**ASTROPARTICLE AND GAMMA-RAY
PHYSICS IN SPACE**

Editors

A. Morselli, P. Picozza

Consorzio Interuniversitario per la Fisica Spaziale
INTERNATIONAL SCHOOL OF SPACE SCIENCE, 2001 Course on
**ASTROPARTICLE AND GAMMA-RAY
PHYSICS IN SPACE**

24

2002

Conorzio Interuniversitario per la Fisica Spaziale

**INTERNATIONAL SCHOOL OF SPACE SCIENCE, 2001 COURSE ON
ASTROPARTICLE AND GAMMA-RAY PHYSICS IN SPACE**

FRASCATI PHYSICS SERIES

Series Editor
Calogero Natoli

Technical Editor
Luigina Invidia

Cover: "A cup of milk" by Silvia Vernetto

Volume XXIV

Istituto Nazionale di Fisica Nucleare – Laboratori Nazionali di Frascati
Divisione Ricerca – SIS – Ufficio Pubblicazioni
P.O. Box 13, I-00044 Frascati (Roma) Italy
email: invidia@lnf.infn.it

White page

FRASCATI PHYSICS SERIES

Conorzio Interuniversitario per la Fisica Spaziale
INTERNATIONAL SCHOOL OF SPACE SCIENCE 2001 COURSE ON
ASTROPARTICLE AND GAMMA-RAY PHYSICS IN SPACE

Copyright © 2002, by INFN Laboratori Nazionali di Frascati
SIS – Ufficio Pubblicazioni

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN 88-86409-30-3

Printed in Italy by Poligrafica Laziale, P.le della Stazione 1, 00044 Frascati

FRASCATI PHYSICS SERIES

Volume XXIV

Conorzio Interuniversitario per la Fisica Spaziale

**INTERNATIONAL SCHOOL OF SPACE SCIENCE, 2001 COURSE ON
ASTROPARTICLE AND GAMMA-RAY PHYSICS IN SPACE**

Editors

A. Morselli, P. Picozza



L'Aquila (Italy), August 30–September 7, 2001

The Director of the School

U. Villante (University of L'Aquila)

The Directors of the Course

G. Barbiellini (University of Trieste)

C. Bonifazi (ASI)

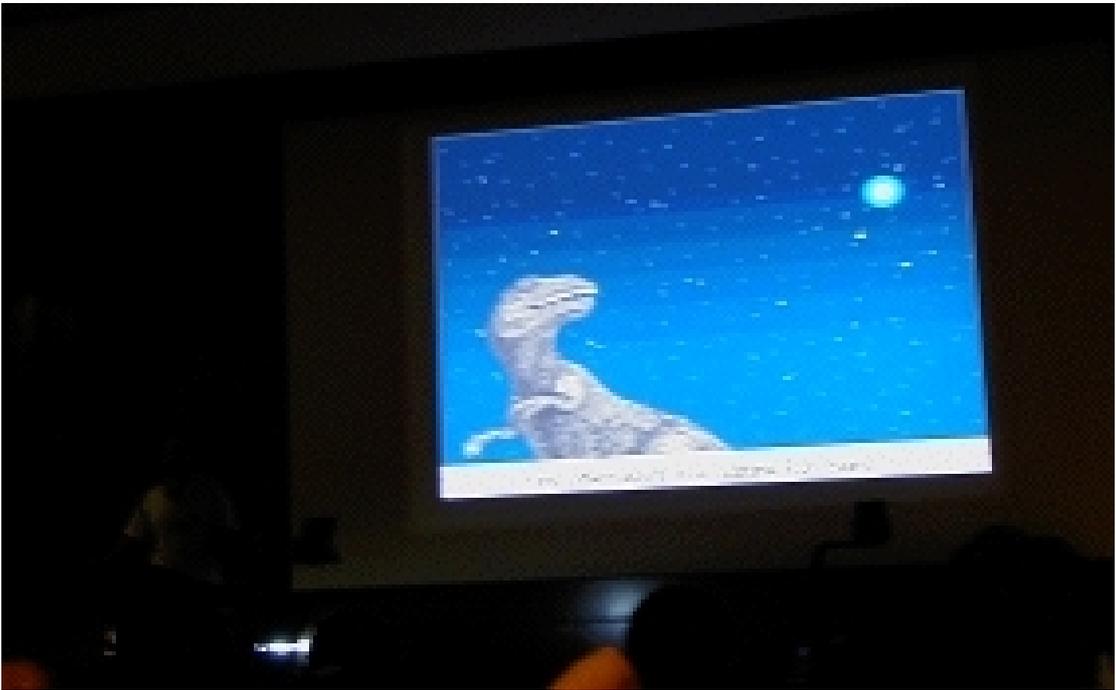
A. Morselli (University of "Tor Vergata and INFN Roma2)

P. Picozza (University of "Tor Vergata and INFN Roma2)

The International School of Space Science is supported by:

INFN - Istituto Nazionale di Fisica Nucleare; ASI - Agenzia Spaziale Italiana; University of Roma "Tor Vergata; University of L'Aquila, Area di Astrogeofisica; Regione Abruzzo; Comune dell'Aquila; Alenia Spazio.





PREFACE

The International School of Space Science (ISSS), has been established since 1990 in L'Aquila (Italy) by the *Conorzio Interuniversitario per la Fisica Spaziale* (CIFS) in cooperation with the Scuola Superiore Guglielmo Reiss Romoli (SSGRR). The aim of the School is the organization of courses, for a limited number of attending young scientist, on the different aspects of the space activities; previous courses addressed respectively :

1992-*Ultraviolet and optical astronomy from space*; 1993-*Solar system plasma physics*; 1994-*X-ray astronomy*; 1995-*Physics of planets and planetary environments*; 1996-*Space Science from the space station*; 1997-*The sun as seen from space*; 1998-*3K cosmology from space*, 1999-*High resolution observation in astronomy*; 2000 *Sun Earth connection and space weather*.

The 2001 Course of the International School of Space Science was held from August 30th to September 7th, 2001 at the Scuola Superiore "Guglielmo Reiss Romoli" near L'Aquila on *Astroparticle and Gamma-ray Physics in Space* and referred to the study of the fundamental laws of the Nature through the observation of the most energetic events in our Universe by the detection of cosmic rays and gamma rays from space. This field is now emerging as a very stimulating and active one and it is complementary to the particle physics done at accelerators, underground, and on mountain laboratories. Its main focus is the study of theoretical items concerning dark matter, cosmic ray origin and propagation, gamma ray sources in the Universe and on the related balloon and space experiments. The vitality of the field is witnessed by the large number of experiments in progress and planned, like the balloon experiments CAPRICE, BESS, ISOMAX, HEAT, the space experiments PAMELA and AMS for antimatter cosmic rays detection, ACE and the planned ACCESS for nuclear composition, NINA, SOHO and SAMPEX for the study of Solar Energetic particles and low energy cosmic ray composition, AGILE and GLAST for gamma rays, EUSO for the extreme high energy cosmic rays. Moreover, since this field needs the use of advanced technologies both on earth and on space, it is a field where advanced research meets advanced technological industries with very fruitful exchange and offers a chance to involve small and medium high-tech enterprises in a joint discussion on the technology exchange between space research and space industries.

The credit for the scientific success of the Course mainly goes to the students, with their active and stimulating participation and to the speakers for their effort in presenting clear and deep talks, for their presence during almost all the period of the school, for their availability to answer questions and to participate in discussions also outside the "canonical" time and for supplying their written contributions within the very tough schedule decided to have a very update reference book. Also we have to thank the Director of the International School of Space Science, Umberto Villante, for all the support and suggestions, all the staff of the Reiss Romoli for their professionalism and kindness, our secretariat staff Gabriella Ardizzoia, Paola Solini and Simona Martana, our sponsors: INFN - Istituto Nazionale di Fisica Nucleare, ASI - Agenzia Spaziale Italiana, University of Roma "Tor Vergata", University of L'Aquila, Area di Astrogeofisica, Regione Abruzzo, Comune dell'Aquila and Alenia Spazio; Liù Catena and Anna Minella for her help in the preparation of the Course, Luigina Invidia and Andrea Lionetto for the help in the editorial procedure.

The book is available in electronic format in
<http://www.roma2.infn.it/inf/aldo/ISSS01.html>

February 2002

The Editors

CONTENTS

Preface
Introduction to astroparticle and gamma ray physics in space
M.S.Longair	Introduction to high energy astrophysics..... 3
J.Ellis	Astroparticle and Cosmology : Learning from cosmic and gamma rays 49
G. Amelino-Camelia	Quantum-Gravity Phenomenology with Gamma Rays and UHE Cosmic Rays 79
A.Balbi	High precision cosmology 95
Cosmic Rays
E. Berezhko	Cosmic Ray Acceleration and High-Energy Gamma-Ray Production in Supernova Remnants 107
M.Simon	The interpretation of radioactive isotopes in the cosmic radiation and the link between the diffusion halo model and the leaky box model 119
F.Donato	Propagation of galactic cosmic ray nuclei and antiprotons in a diffusion model 137
P.Spillantini	Balloon experiments..... 149
M.Boezio	Analysis of cosmic-ray data. Examples with antiprotons and positrons 173
Y. Butt	Supernova Remnants as Hadronic Cosmic Ray Accelerators?... 183
M.Casolino	Radiation in Space: Causes Measures and Countermeasures. ... 195
G.Esposito	Protons and Lepton Trapped in the Earth's Radiation Belts 217
R.Sparvoli	Cosmic rays in the Earth vicinity. 223
P.Spillantini	The antimatter component of cosmic rays and the PAMELA experiment 249
R.Battiston	The ALPHA Magnetic Spectrometer, a Particle Physics Experiment in Space 261
L.Scarsi	EUSO: Using High Energy Cosmic Rays and Neutrinos as Messengers from the Unknow Universe 279
C. Sbarra	Photomultiplers for the New TOF System of AMS-02 Space Experiment..... 297

The gamma-ray side of the Universe	
P.Fleury	Gamma Astronomy - Results from Ground Observations 305
G.Barbiellini	Science and Technology: The AGILE Scientific Instrument 325
M.Tavani	Gamma -ray astrophysics and the AGILE mission 333
R.Bellazzini	The GLAST Gamma ray large area space telescope 353
A.Morselli	Gamma ray astroparticle physics with GLAST 363
R.Battiston	AMS- γ : High Energy Photon Detection with the ALPHA Magnetic Spectrometer on the ISS..... 381
A.Celotti	The physics of Active Galactic Nuclei 391
A. De Angelis	The Simulation of Astroparticle Experiments and the GEANT4 Toolkit 399
G.Kanbach	MEGA - a new telescope for medium energy gamma-ray astronomy 409
S. Rebecchi	An introduction to ASDC a new ASI facility 417
V. Cocco	Geant Simulation of the AGILE Gamma-Ray Imaging Detector..... 421
C. Pittori	Presentation of the AGILE Reconstruction Method and Kalman Filter Algorithms for Gamma-Ray Silicon Detectors in Space .. 425
The dark side of the Universe	
P. De Bernardis	The density of the Universe 433
J.Primack	The Nature of Dark Matter 449
P.Ullio	Exotic cosmic rays from pair annihilations of dark matter wimps 475
A.Lionetto	MSSM parameters space regions allowed by indirect neutralino detection 485
The engines of the Universe	
M.Vietri	What have we learned about gamma ray bursts from afterglows? 493
A.Dar	The threat to life from Eta Carinae and gamma-ray bursts 513
F.Longo	Gamma-ray satellites and gamma-ray bursts 525
N.Omodei	Emission Model and GRB Simulations..... 531
C. Aramo	AUGER observatory: the world's largest cosmic ray detector .. 539
A.Bettini	(New) Neutrino Physics at Gran Sasso Laboratory 545
J.Primack	Probing galaxy formation with high energy gamma-rays 559
Participants 583

Introduction to Astroparticle and Gamma Ray Physic in Space

M.S.Longair	Introduction to high energy astrophysics
J.Ellis	Astroparticle and Cosmology : Learning from cosmic and gamma rays
G. Amelino-Camelia	Quantum-Gravity Phenomenology with Gamma Rays and UHE Cosmic Rays
A.Balbi	High precision cosmology

INTRODUCTION TO HIGH ENERGY ASTROPHYSICS

Malcolm S. Longair
Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE

ABSTRACT

This introduction to high energy astrophysics is on two parts. In the first, the astrophysical framework within which high energy astrophysical phenomena are observed is described, as well as a number of important astrophysical processes associated with black holes, accretion, hot intracluster gas and jets in active galaxies. In the second, many of the most important radiation processes are described, including the radiation of accelerated electrons, bremsstrahlung, synchrotron radiation, inverse Compton scattering, synchro-Compton radiation, γ -ray processes and relativistic beaming. The emphasis is upon clarity in the description of the basic physics.

1 The Scope of High Energy Astrophysics

High energy astrophysics is a vast subject and this brief review can only scratch the surface of this dynamic and extensive field. I have divided the article into

two parts, the first concerning the *Astrophysical Framework* for high energy astrophysics, and the second *Radiation Processes*. These should provide some of the essential background needed to understand the many excellent papers presented at this School.

I interpret the term *High Energy Astrophysics* to mean a number of things. It certainly includes:

- The astrophysics of non-thermal processes, specifically, the physics of relativistic and non-relativistic protons, nuclei and electrons with power-law energy spectra $N(E) \propto E^{-\gamma} dE$ in a huge variety of different astrophysical environments, and the processes by which the particles were accelerated with this form of spectrum.
- The physics of high temperature plasmas.
- Astrophysics in extreme astrophysical environments.

More details of my approach to these and many others topics can be found in my books *High Energy Astrophysics. Vol. 1 and 2* (1997) (hereafter *HEA1* and *HEA2*), as well as my review in the volume *X-ray Spectroscopy in Astrophysics* (1999) (hereafter *XSA*). I will place the emphasis upon the physics of the processes involved.

Part 1 The Astrophysical Framework

2 White Dwarfs, Neutron Stars and Black Holes

Much of the discussion of high energy astrophysical phenomena centres around the discussion of compact objects. The most compact stars are white dwarfs, neutron stars and black holes. Their stability criteria can be found from the *virial theorem for degenerate stars*. In the case of stellar material with a non-relativistic, non-degenerate equation of state, the condition for stability is given by the virial theorem in the form $3(\gamma - 1)U + \Omega = 0$, where U is the total internal thermal energy, γ is the ratio of specific heats of the stellar material and $\Omega = GM^2/R$ is the gravitational potential energy of the star. M is its mass and R a suitably chosen radius.

An important extension of these results is to the case of degenerate stars. The pressure is associated with the fact that fermions such as electrons, protons and neutrons cannot occupy the same quantum mechanical state. Simple

arguments show that, in both the relativistic and non-relativistic cases, the pressure is independent of temperature:

$$\text{Non-relativistic : } p \approx \frac{\hbar^2}{m_e} \left(\frac{\rho}{m_p} \right)^{5/3}, \quad \text{Relativistic : } p \approx \hbar c \left(\frac{\rho}{m_p} \right)^{4/3}.$$

Inserting the non-relativistic formula into the equation of hydrostatic support enables solutions for low-mass white dwarfs and neutron stars to be found.

In the *relativistic case*, the total internal energy U is

$$U = V\varepsilon = 3Vp \approx V\hbar c \left(\frac{\rho}{m_p} \right)^{4/3}, \quad (1)$$

where ε is the energy density of the gas, which in the relativistic limit is related to the pressure by $p = \frac{1}{3}\varepsilon$. Therefore, the virial theorem becomes:

$$2U = |\Omega|, \quad 2V\hbar c \left(\frac{\rho}{m_p} \right)^{4/3} = \frac{1}{2} \frac{GM^2}{R}. \quad (2)$$

We can now approximate $V \approx R^3$ and $\rho V = M$. Therefore, (2) becomes

$$\frac{2\hbar c}{R} \left(\frac{M}{m_p} \right)^{4/3} = \frac{1}{2} \frac{GM^2}{R}. \quad (3)$$

This is the important result. Both the right and left-hand sides of the expression depend upon radius R in the same way. Dropping constants of order unity, the mass of the star is:

$$M \approx \frac{1}{m_p^2} \left(\frac{\hbar c}{G} \right)^{3/2} \approx 2M_\odot. \quad (4)$$

This is the famous expression for the *Chandrasekhar mass* and is the upper limit for the mass of relativistic degenerate stars. If the mass were any greater, the gravitational term would always dominate and, since both the pressure and gravitational terms depend upon mass in the same way, the star would collapse indefinitely. The exact expression is

$$M_{\text{Ch}} = \frac{5.836}{\mu_e^2} M_\odot, \quad (5)$$

where $\mu_e m_u$ is the average particle mass per electron, which is 2 for a relativistic degenerate gas, except for hydrogen. A similar result applies for neutron stars,

but the exact result is more difficult to work out since the effects of general relativity cannot be ignored.

It is instructive to write the Chandrasekhar mass in terms of the *gravitational fine-structure constant*. The usual fine-structure constant is $\alpha = e^2/4\pi\epsilon_0\hbar c \approx 1/137$. The equivalent for gravity can be found by replacing $e^2/4\pi\epsilon_0$ by Gm_p^2 .

$$\alpha_G = \frac{Gm_p^2}{\hbar c} \approx 5.6 \times 10^{-39}, \quad (6)$$

reflecting the different strengths of the electrostatic and gravitational forces. Therefore the Chandrasekhar mass is $M \approx m_p\alpha_G^{-3/2}$. In terms of fundamental constants, stars are objects which possess about 10^{58} protons.

If the mass of the compact star is greater than about $2M_\odot$, it is inevitable that the object collapses to a *black hole*. It is convenient to consider separately the spherically symmetric, non-rotating black holes and those with non-zero angular momentum.

- **Schwarzschild Black Holes** are non-rotating and spherically symmetric. There is a surface of infinite redshift at the *Schwarzschild radius* R_g .

$$R_g = \frac{2GM}{c^2} = 3 \left(\frac{M}{M_\odot} \right) \text{ km}. \quad (7)$$

For an observer at infinity, radiation emitted at this radius is redshifted to zero frequency – this is why the black holes are called ‘black’. There is a last stable circular orbit about the black hole at radius $r = 3R_g$. If matter were placed on a circular orbit at radial distance less than $3R_g$, it would rapidly spiral into the black hole - this is why the black holes are called ‘holes’.

- **Kerr Black Holes** possess angular momentum J . The surface of infinite redshift occurs at radius

$$R_+ = \frac{GM}{c^2} + \left[\left(\frac{GM}{c^2} \right)^2 - \left(\frac{J}{Mc} \right)^2 \right]^{1/2}. \quad (8)$$

For a maximally rotating Kerr black hole, the term in square brackets is zero and then

$$R_+ = \frac{GM}{c^2}, \quad J_{\max} = \frac{GM^2}{c}. \quad (9)$$

The last stable circular orbit depends upon whether the test particle is corotating or counter-rotating with respect to the black hole:

$$\text{Corotating} \quad r = \frac{GM}{c^2} \quad : \quad \text{Counter-rotating} \quad r = \frac{9GM}{c^2}. \quad (10)$$

The corotating case in (10) is very important since it demonstrates that matter can approach much closer to $r = 0$ for a maximally rotating Kerr black hole than it can for a Schwarzschild black hole.

3 Accretion onto Compact Objects

Accretion of matter under gravity onto compact objects is a very powerful source of energy. Consider a mass m falling from infinity onto a compact mass M which has radius R . The kinetic energy of collapse is dissipated as heat when it hits the surface of the star.

$$\frac{GMm}{R} = \frac{1}{2}mv^2 = \text{Heat Dissipated}. \quad (11)$$

Suppose the matter is accreted at a rate \dot{m} . Then, the rate of dissipation of energy is

$$\frac{dE}{dt} = \frac{GM\dot{m}}{R}. \quad (12)$$

Expressing this result in terms of the *Schwarzschild radius*,

$$\frac{dE}{dt} = \frac{\dot{m}c^2}{2} \left(\frac{R_g}{R} \right). \quad (13)$$

For a neutron star of mass $1 M_\odot$, $R = 10$ km, according to this estimate, about 1/6 of the rest mass energy of the infalling matter is dissipated at heat. This is an overestimate since this calculation neglects the effects of general relativity which cannot be ignored in such compact objects. It is convenient to write the rate of energy generation, or the luminosity, of the source as

$$\frac{dE}{dt} = \eta \dot{m}c^2, \quad (14)$$

where η describes the efficiency of the accretion process. The following table shows the maximum efficiencies obtainable by various astrophysical processes.

Efficiency of Energy Production: Energy Released = ηmc^2

<i>Form of energy production</i>	η
Chemical energy	10^{-9}
Nuclear energy	10^{-2}
Schwarzschild black holes	0.06
Kerr black holes	0.43
Rotational energy of black hole	0.29

- Thus, the accretion of matter onto rotating black holes is the most powerful energy source for objects such as X-ray binaries and active galactic nuclei.
- Black holes are the most compact objects which a mass M can have and so the shortest time-scale which can be associated with an object of mass M is $t \sim R_g/c \approx 10^{-5}(M/M_\odot)$ s.

The amount of energy which can be obtained from accretion is not, however, unlimited. The maximum luminosity L is set by the balance between the attractive force of gravity acting on the protons and the repulsive force of radiation pressure acting on the electrons. In a fully ionised plasma, the protons and electrons are strongly coupled by electrostatic forces between protons and electrons:

$$f_{\text{grav}} = \frac{GMm_p}{r^2}, \quad f_{\text{rad}} = \sigma_T N_{\text{ph}} \frac{h\nu}{c},$$

where $N_{\text{ph}} = L/4\pi r^2 c$ is the flux density of photons at distance r from the source and σ_T is the Thomson cross-section. Thus, the attractive and repulsive forces both depend upon r^{-2} . Equating these forces, the *Eddington luminosity* L_E is

$$L_E = \frac{4\pi GMm_p c}{\sigma_T} = 1.3 \times 10^{31} \left(\frac{M}{M_\odot} \right). \quad (15)$$

The Eddington limit depends only upon the mass of the star or black hole and applies to steady accretion. There are various ways by which super-Eddington luminosities can be obtained, but (15) is a useful guideline for the maximum luminosity which can be liberated in the steady state.

4 Accretion Discs

How is the potential energy of the matter onto the black hole released? It is most unlikely that the matter has zero angular momentum, because of random gravitational perturbations acting on the infalling matter. The matter then collapses along the angular momentum axis of the infalling material. The same process occurs in the formation of stars.

In the case of a black hole, a disc forms about the black hole and extends in to the last stable orbit. To do this, the matter must transfer its angular momentum outwards through the disc through the action of viscous forces, which at the same time results the material of the disc is heated by the viscous forces. These models for accretion discs have been the subject of a great deal of study.

Evidence for this picture has been found in the X-ray spectra of the nuclei of certain Seyfert 1 galaxies. There is convincing evidence from spectral observations with the ASCA X-ray observatory for the presence of fluorescent $K\alpha$ emission at 6.4 keV in their spectra. The line is broadened on its low energy wing. The effect has been observed with reasonable signal-to-noise ratio in the spectrum of the Seyfert galaxy MCG-6-30-15. The intriguing aspect of these observations is that the redshift of the line extends to about 4.5 keV. The preferred model to account for this large redshift is that it is associated with the combined gravitational and transverse Doppler shifts associated with matter close to the last stable orbit about the black hole. For a Schwarzschild black hole, the rotational speed of matter at the last stable orbit is $c/2$ and that the maximum combined gravitational plus Doppler shift is $\nu_{\text{obs}} = \nu/\sqrt{2}$. If this interpretation is correct, it is of the greatest importance since it would mean that we can observe matter very close to the last stable orbit about the black hole and so have the opportunity to study the behaviour of matter in strong gravitational fields.

5 The Masses of Black Holes in X-ray binaries and the Nuclei of Galaxies

The methods of classical astrometry can be used to estimate the masses of the invisible compact objects which are the source of the intense X-rays observed in X-ray binary systems. In those systems in which the compact companion

is undoubtedly a rapidly rotating neutron star, their masses always turn out to be about $1.4 M_{\odot}$. In some of the cases in which there is no evidence for a neutron star, the masses of the unseen objects are significantly greater than $3 M_{\odot}$, indicating that the objects must be stellar mass black holes. In some cases, the compact object has mass about $10 M_{\odot}$. The evidence is compelling that these are indeed stellar mass black holes and they are of special importance since they are the closest black holes available for detailed study.

In the past, there was circumstantial evidence that supermassive black holes are the sources of energy for the most extreme active galactic nuclei such as the quasars and the nuclei of Seyfert galaxies. Nowadays, much more care has to be taken in establishing that there are massive point-like objects in the nuclei of galaxies. The key diagnostic is the observation that the test objects move in circular Keplerian orbits about the nucleus. This has demonstrated by the HST for M87, the nearby giant elliptical with a prominent jet.

The best extragalactic example has made use of the H_2O masers observed by VLBI in the nucleus of the nearby galaxy NGC 4258. In this case, the rotational speeds of the masers follow a Keplerian relation as far in as they can be measured. Even better are recent estimates of the mass of the black hole in the centre of our Galaxy by following the motions of stars observed at $2 \mu\text{m}$ (Ghez et al 2000). In this case, the acceleration vectors of some of the prominent infrared stars in the Galactic Centre have been measured by very precise astrometry and these intersect at the location of the non-thermal source Sgr A*, which has long been suspected to be at the centre of the Galaxy. A summary of the masses and mass densities found in the nuclei of selected galaxies are listed below.

Galaxy	Mass of Nucleus (M_{\odot})	Mass Density ($M_{\odot} \text{pc}^{-3}$)
NGC 4258	3.6×10^7	$\geq 4 \times 10^9$
M87	10^9	$\geq 10^6$
NGC3115	10^6	$\geq 10^8$
Galactic Centre	2×10^6	$\geq 2 \times 10^{12}$
Globular Cluster	10^6	$\leq 10^5$

In NGC 4258, the mass density is at least 40,000 times greater than that of the densest globular clusters in our Galaxy. If the mass consisted of stars, the separation between them would be only about 100 AU and their collision times

less than 10^8 years. Similar arguments can be made about the nature of the object present in the centre of our Galaxy – it is wholly convincing that there is a black hole with mass about $2 \times 10^6 M_\odot$ lurking in the centre of our Galaxy.

6 Clusters of Galaxies

Clusters of galaxies are the largest gravitationally-bound systems in the Universe and are often the seat of high-energy astrophysical activity. Of special importance is the fact that the intergalactic gas within clusters of galaxies is not only very hot, and hence a luminous source of X-rays, but also enables the mass of the cluster and the gaseous environment to be determined in some detail. This has become a particularly important area with the spectacular results provided by the Chandra and Newton-XMM satellites. The calculation proceeds as follows.

The equation of hydrostatic support for the intergalactic gas in a cluster of galaxies is

$$\frac{dp}{dr} = -\frac{GM(< r)\rho}{r^2}. \quad (16)$$

The pressure is related to the local gas density ρ through the equation of state $p = \rho kT / \mu m_H$ where m_H is the mass of the hydrogen atom and μ is the mean molecular weight of the gas. Differentiating the equation of state with respect to r and substituting into (16), we find

$$M(< r) = -\frac{kTr^2}{G\mu m_H} \left[\frac{d(\log \rho)}{dr} + \frac{d(\log T)}{dr} \right]. \quad (17)$$

Thus, the mass distribution within the cluster can be determined if the variation of the gas density and temperature with radius are known. We will show in Sect. 9.2 how the bremsstrahlung spectrum of the intracluster gas can be used to determine both its temperature and density distribution. In some clusters, the gas is dense enough to cool over cosmological time-scales resulting in *cooling flows* (see Sect. 9.3).

Note that the hot gas acts as a tracer of the *total gravitating mass* in the cluster, including the dark matter which defines the gravitational potential. Böhringer (1995) gives a nice example of his results on the Perseus cluster of galaxies. The typical total masses of rich clusters lie in the range 5×10^{14} to $5 \times 10^{15} M_\odot$, of which about 5% is attributable to the mass contained in

the visible parts of galaxies and about 10 to 30% to hot intracluster gas. The remaining 60 to 85% of the mass is in some form of dark matter. Typically, the iron abundance of the hot gas is between about 20 and 50% of the solar value, indicating that the intergalactic gas has been enriched by the products of stellar nucleosynthesis.

The Chandra observations show that, while some clusters such as the Coma cluster can be well approximated by smooth X-ray profiles, other well-studied clusters display considerable structure in their X-ray distributions. Some of these are certainly due to the presence of powerful radio sources in the clusters. Presumably these events compress the gas allowing it to radiate with greater intensity. These considerations lead another characteristic of active galactic nuclei, the presence of jets.

7 Jets

One of the remarkable phenomena which seem to accompany most manifestations of high energy astrophysical processes associated with compact objects is the presence of highly collimated jets of relativistic material. The best known of these are those associated with extragalactic radio sources and objects such as M87 and 3C273. Some of the more important aspects of their role in astrophysics can be summarised as follows:

- The most common examples of relativistic jets are found in the extragalactic radio sources. In sources such as Cygnus A, highly collimated jets of relativistic particles emerge from the nucleus of the galaxy in opposite directions and inflate huge radio emitting cocoons. The locations at which the jets encounter the surrounding interstellar and intergalactic gas result in 'hot-spots' where particle acceleration takes place. These particles inflate the cocoon as the jet continues to penetrate the surrounding medium. The material of the jet must travel at speeds close to the speed of light and the hot-spots themselves advance through the surrounding medium at speeds up to about 0.1 or 0.2 c .
- The jets can disrupt cool clouds which they encounter as they pass out through the galaxy. We have found spectacular examples of this phenomenon in the case of the most powerful double radio sources at redshifts $z \sim 1$ (Best et al. 2000). In the *alignment effect*, intense optical

emission is observed aligned with the radio jets. Our observations have shown that, in the smaller double radio sources, the excitation mechanism is strong shocks caused by the interaction between the jets and the cool clouds. The origin of the clouds is not established, but they may well be associated with cooling flows in the surrounding intergalactic gas. In the case of sources such as Cygnus A and Perseus A, there is direct evidence for the hot gas associated with the central galaxy being compressed and buffeted by the cocoons of the radio sources.

- The radio galaxies and radio quasars found in complete samples are in remarkable agreement with orientation-based *unification schemes*. The idea behind unification is that what one observes in active galaxies is strongly influenced by the angle at which the observer is located with respect to the axis of ejection of the jet. If there is an obscuring torus about the active nucleus, the quasar-like activity in the nucleus is only observed if the observer lies within a certain angle with respect to the axis of the torus. This scheme can account for the properties of the radio galaxies and radio quasars found in, for example, the 3CR sample of radio sources if the opening angle of the torus is about 45° . Similar schemes can account for the differences between the Seyfert I and Seyfert II galaxies – in the former, the nuclear regions are observed directly, whereas in the latter, these regions can only be observed in the reflected light of clouds well outside the region of the torus (see Antonucci 1993).
- One of the big surprises of the study of luminous compact radio sources was the discovery of superluminal motions in sources such as 3C273 and 3C120. In 3C273, the jets are observed to move out from the nucleus at speeds up to 8 to 10 times the speed of light. This phenomenon has been found in many of the compact radio sources found in high frequency surveys. In the favoured relativistic ballistic model, it is assumed that this is an optical illusion associated with the fact that the jet is approaching the observer at a speed close to the speed of light at an angle close to the line of sight (see Sect. 14). Kellermann (2001) pointed out that there is a strong bias in favour of observing the largest superluminal velocities, because the large blue-shifts of the approaching source components greatly amplify their observed flux densities. When account is taken of this bias,

the average observed value is probably closer to $v_{\text{obs}} = 2c$. Relativistic velocities have been observed in the components of the optical jet of M87.

- Further evidence for relativistic beaming has been found in the extreme γ -ray luminosities of the quasars detected by the Compton Gamma-Ray Observatory. These are so extreme in luminosity and variability, that the effects of photon-photon scattering would destroy the γ -ray photons within the source region. Fortunately, it is also found that the majority of these extreme γ -ray sources are also associated with objects which exhibit superluminal motion of their radio source components. The natural explanation for these enormous luminosities is that they are boosted by the same type of relativistic beaming which is responsible for the observation of superluminal motions.
- To everyone's surprise, in 1994 Mirabel and Rodriguez discovered that one of the Galactic X-ray binary source GRS 1915+105, which contains a stellar mass black hole, exhibits superluminal motion. They demonstrated convincingly that this class of source provides more or less exact counterparts for the phenomena observed in radio galaxies and radio quasars, but scaled down by a factor of about one million - they named these objects *microquasars*. The reason for this rather precise scaling is that both the luminosity and the time-scale of variability of the sources are proportional to the mass of the black hole. The only significant difference is that the microquasars are powered by accretion due to the dragging of mass off the primary star of the binary, whereas in the extragalactic case, the mass is presumed to originate by infall and dissipation from the interstellar gas in the parent galaxy. The scaling down by a factor of one million means that phenomena, which would only be observable over periods of decades, or centuries, in the extragalactic case can be observed within minutes or hours.
- The last example of sources in which relativistic beaming must play an important role are the *γ -ray bursts*. It is now certain that these objects lie at cosmological distances and are the most luminous objects in the Universe for the seconds or minutes during which the burst lasts. To account for the luminosities and short duration of the bursts, relativistic motions are again invoked, this time in the form of a relativistic blast

wave resulting from a catastrophic event, probably associated with the collapse of the core of a massive star, or possibly with the coalescence of a binary neutron star system.

Part 2 Radiation Processes

The objective of these notes is to indicate how essentially all the important features of the radiation processes encountered in high energy astrophysics can be understood in terms of the elaboration of a few simple basic principles. The processes to be considered below include: the fundamentals of the radiation of charged particles, bremsstrahlung, synchrotron radiation, inverse Compton radiation, synchro-Compton radiation, γ -ray processes and relativistic beaming. The acceleration of charged particles is splendidly reviewed in this volume by Professor Berezhko. The books which I find most useful in understanding these processes are those by Rybicki and Lightman (1979), Jackson (1975) and Blandford, Netzer and Woltjer (1990), as well as my own versions of the processes (Longair 1997, 1999). This part is a highly abridged version of my review in *XSA*.

8 Basic Radiation Concepts

The key relation is the radiation loss rate of an accelerated charged particle in the non-relativistic limit

$$-\left(\frac{dE}{dt}\right)_{\text{rad}} = \frac{|\ddot{\mathbf{p}}|^2}{6\pi\epsilon_0 c^3} = \frac{q^2 |\ddot{\mathbf{r}}|^2}{6\pi\epsilon_0 c^3}. \quad (18)$$

$\mathbf{p} = q\mathbf{r}$ is the *dipole moment* of the accelerated electron with respect to some origin. This formula is often referred to as the radiation loss rate for *dipole radiation*. I will use *SI units* throughout this survey. In the non-relativistic limit, the radiation loss rate depends only upon the acceleration of the charged particle. This formula can be derived using elementary concepts in electromagnetism (see *HEA1* or *XSA*).

The strength of the electric field in the radiated wave at angle θ to the acceleration vector is

$$E_\theta = \frac{q\ddot{r} \sin\theta}{4\pi\epsilon_0 c^2 r}. \quad (19)$$

Correspondingly, the rate loss of energy through the solid angle $d\Omega$ at distance r from the charge is

$$-\left(\frac{dE}{dt}\right)_{\text{rad}} d\Omega = \frac{|\ddot{\mathbf{p}}|^2 \sin^2 \theta}{16\pi^2 Z_0 \varepsilon_0^2 c^4 r^2} r^2 d\Omega = \frac{|\ddot{\mathbf{p}}|^2 \sin^2 \theta}{16\pi^2 \varepsilon_0 c^3} d\Omega. \quad (20)$$

To find the total radiation rate, we integrate over all solid angles to find the result (18) which is sometimes called *Larmor's formula*. These formulae embody the three essential properties of the radiation of an accelerated charged particle.

1. The total radiation rate is given by (18). In this formula, the acceleration is the *proper acceleration* of the charged particle and the radiation loss rate is measured in the instantaneous rest frame of the particle.
2. The *polar diagram* of the radiation is of *dipolar* form, that is, the electric field strength varies as $\sin \theta$ and the power radiated per unit solid angle varies as $\sin^2 \theta$. There is no radiation along the acceleration vector and the field strength is greatest at right angles to the acceleration vector.
3. The radiation is *polarised* with the electric field vector lying in the direction of the acceleration vector, as projected onto a sphere at distance r from the charged particle.

These are very useful rules which enable us to understand the radiation properties of particles in different astrophysical situations. These rules are applicable in the *instantaneous rest frame* of the particle and so we need to look carefully at what an external observer sees when the particle is moving at a relativistic velocity.

8.1 Thomson Scattering

Let us apply these results to the case of *Thomson scattering*, the scattering of electromagnetic waves by free electrons in the classical limit. The acceleration of the electron is due to the E -fields of the incident electromagnetic waves. If α is the angle between the direction of the incident flux of radiation and the direction of observer, the loss rate is

$$-\left(\frac{dE}{dt}\right) d\Omega = \frac{e^4}{16\pi^2 m_e^2 \varepsilon_0^2 c^4} (1 + \cos^2 \alpha) \frac{S}{2} d\Omega \quad (21)$$

where S is incident flux density of unpolarised radiation. The scattered intensity can be written in terms of a differential scattering cross-section $d\sigma_{\text{T}}$

$$\frac{d\sigma_{\text{T}}(\alpha)}{d\Omega} = \frac{\text{energy radiated per unit time per unit solid angle}}{\text{incident energy per unit time per unit area}}. \quad (22)$$

The differential cross-section for Thomson scattering is therefore

$$d\sigma_{\text{T}}(\alpha) = \frac{e^4}{16\pi^2\epsilon_0^2 m_e^2 c^4} \frac{(1 + \cos^2 \alpha)}{2} d\Omega. \quad (23)$$

In terms of the *classical electron radius* $r_e = e^2/4\pi\epsilon_0 m_e c^2$, this can be expressed

$$d\sigma_{\text{T}} = \frac{r_e^2}{2} (1 + \cos^2 \alpha) d\Omega. \quad (24)$$

To find the total cross-section for scattering, we integrate over all angles α ,

$$\sigma_{\text{T}} = \int_0^\pi \frac{r_e^2}{2} (1 + \cos^2 \alpha) 2\pi \sin \alpha d\alpha = \frac{8\pi r_e^2}{3} = 6.653 \times 10^{-29} \text{ m}^2. \quad (25)$$

This is Thomson's famous result for the total cross-section for scattering by stationary free electrons, the *Thomson cross-section*. It appears in many radiation formulae.

A distinctive feature of the process is that the scattered radiation is polarised, even if the incident beam is unpolarised. The degree of polarisation can be found from a simple calculation:

$$\Pi = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}} = \frac{1 - \cos^2 \alpha}{1 + \cos^2 \alpha}. \quad (26)$$

8.2 A useful relativistic invariant

Very often we have to transform the energy loss rate by radiation, dE/dt , from one inertial frame of reference to another – it turns out that dE/dt is a Lorentz invariant between inertial frames. To the expert in relativity, this is obvious. The total energy emitted in the form of radiation is the ‘time’ component of the momentum four-vector $[E/c^2, \mathbf{p}]$ and dt is the time component of the displacement four-vector $[dt, d\mathbf{r}]$. Therefore, both the energy dE and the time interval dt transform in the same way between inertial frames of reference and their ratio dE/dt is also an invariant, that is,

$$\frac{dE}{dt} = \frac{dE'}{dt'}. \quad (27)$$

We can derive from this expression the radiation rate as observed by the external observer who measures the velocity and acceleration of the electron to be \mathbf{a} and \mathbf{v} respectively, the proper acceleration measured in the instantaneous rest frame of the electron being \mathbf{a}_0 . To relate \mathbf{a}_0 , \mathbf{a} and \mathbf{v} , it is simplest to equate the norms of the four-accelerations of the accelerated electron in the frames S and S'.

$$\left(\frac{dE}{dt}\right)_{\text{in S}} = \frac{e^2\gamma^4}{6\pi\epsilon_0c^3} \left[\mathbf{a}^2 + \gamma^2 \left(\frac{\mathbf{v} \cdot \mathbf{a}}{c}\right)^2 \right]. \quad (28)$$

Another useful exercise is to resolve \mathbf{a} parallel and perpendicular to \mathbf{v} so that

$$\left(\frac{dE}{dt}\right)_{\text{in S}} = \frac{e^2\gamma^4}{6\pi\epsilon_0c^3} (|a_\perp|^2 + \gamma^2|a_\parallel|^2). \quad (29)$$

8.3 Parseval's theorem and the spectral distribution of the radiation of an accelerated electron

Parseval's theorem provides an elegant method of relating the kinematic history of the particle to its radiation spectrum. We introduce the Fourier transform of the acceleration of the particle through the Fourier transform pair:

$$\dot{\mathbf{v}}(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \dot{\mathbf{v}}(\omega) \exp(-i\omega t) d\omega \quad (30)$$

$$\dot{\mathbf{v}}(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \dot{\mathbf{v}}(t) \exp(i\omega t) dt. \quad (31)$$

Parseval's theorem states that $\dot{\mathbf{v}}(\omega)$ and $\dot{\mathbf{v}}(t)$ are related by the integrals:

$$\int_{-\infty}^{\infty} |\dot{\mathbf{v}}(\omega)|^2 d\omega = \int_{-\infty}^{\infty} |\dot{\mathbf{v}}(t)|^2 dt. \quad (32)$$

It is then straightforward to show that

$$I(\omega) = \frac{e^2}{3\pi\epsilon_0c^3} |\dot{\mathbf{v}}(\omega)|^2. \quad (33)$$

where $I(\omega)$ is the total energy per unit bandwidth emitted throughout the period during which the particle is accelerated. For a distribution of particles, (33) must be integrated over all the particles contributing to the radiation at frequency ω .

9 Bremsstrahlung

Bremsstrahlung is the radiation associated with the acceleration of electrons in the electrostatic fields of ions and the nuclei of atoms.

9.1 Encounters between Charged Particles

To begin with, let us study the collision of a high energy proton or nucleus with the electrons of a fully ionised plasma. The charge of the high energy particle is ze , its mass M , and it is assumed that it is undeviated in the encounter with the electron; b , the distance of closest approach of the particle to the electron, is called the *collision parameter* of the interaction. The total *momentum impulse* given to the electron in the encounter is $p = \int F dt = ze^2/2\pi\epsilon_0bv$ and the kinetic energy transferred to the electron is

$$\frac{p^2}{2m_e} = \frac{z^2 e^4}{8\pi^2 \epsilon_0^2 b^2 v^2 m_e} = \text{energy lost by high energy particle.} \quad (34)$$

The *average energy loss per unit length* is found by integrating over collision parameters:

$$-\frac{dE}{dx} = \frac{z^2 e^4 N_e}{4\pi \epsilon_0^2 v^2 m_e} \ln \left(\frac{b_{\max}}{b_{\min}} \right). \quad (35)$$

where N_e is the number density of electrons. You may well ask, ‘Why introduce the limits b_{\max} and b_{\min} , rather than work out the answer properly?’ Our approximate methods give rather good answers because the limits b_{\max} and b_{\min} only appear inside the logarithm and hence need not be known very precisely.

This is the simplest example of the type of calculation which is carried out in working out energy transfers and accelerations of electrons and protons in fully ionised plasmas. The logarithmic term $\ln(b_{\max}/b_{\min})$ appears in the guise of what are often referred to as *Gaunt factors* and care has to be taken to use the correct values of b_{\max} and b_{\min} in working out, for example, the radiation spectrum of bremsstrahlung and the electrical conductivity of a plasma.

9.2 The Spectrum and Energy Loss Rate of Bremsstrahlung

In the classical limit, bremsstrahlung is the emission of an electron in an electrostatic encounter with a nucleus. Electrons lose more energy in electron-electron collisions, but these do not result in the emission of dipole radiation since there

is no net electric dipole moment associated with these encounters. First, we need an expression for the acceleration of an electron in the electrostatic field of the nucleus. Then, we take the Fourier transform of the acceleration of the electron and use Parseval's theorem to work out the spectrum of the emitted radiation. Next, we integrate this result over all collision parameters and we have to worry about suitable limits for b_{\max} and b_{\min} .

The relativistic and non-relativistic cases begin in the same way. The accelerations along the trajectory of the electron, a_{\parallel} , and perpendicular to it, a_{\perp} , in its rest-frame are given by

$$a_{\parallel} = \dot{v}_x = -\frac{eE_x}{m_e} = \frac{\gamma Ze^2 vt}{4\pi\epsilon_0 m_e [b^2 + (\gamma vt)^2]^{3/2}} \quad (36)$$

$$a_{\perp} = \dot{v}_z = -\frac{eE_z}{m_e} = \frac{\gamma Ze^2 b}{4\pi\epsilon_0 m_e [b^2 + (\gamma vt)^2]^{3/2}} \quad (37)$$

where Ze is the charge of the nucleus (see *HEA1*).

After some algebra we can find the intensity spectrum of the radiation associated with both accelerations. It turns out that the impulse perpendicular to the direction of travel contributes the greater intensity, even in the non-relativistic case $\gamma = 1$ and results in significant radiation at low frequencies. At high frequencies, there is an exponential cut-off. This can be understood as follows. The duration of the relativistic collision is roughly $\tau = 2b/\gamma v$. Therefore, the dominant Fourier components in the radiation spectrum correspond to frequencies $\nu \approx 1/\tau = \gamma v/2b$ and hence to $\omega \approx \pi v \gamma/2b$, that is, $\omega b/\gamma v \sim 1$. The exponential cut-off tells us that there is little power emitted at frequencies greater than $\omega \approx \gamma v/b$.

The low frequency spectrum has the form

$$I(\omega) = \frac{Z^2 e^6}{24\pi^4 \epsilon_0^3 c^3 m_e^2} \frac{1}{b^2 v^2} \left[1 - \frac{1}{\gamma^2} \left(\frac{\omega b}{\gamma v} \right)^2 \ln^2 \left(\frac{\omega b}{\gamma v} \right) \right]. \quad (38)$$

In the low frequency limit, $\omega b/\gamma v \ll 1$, the second term in square brackets can be neglected and hence

$$I(\omega) = \frac{Z^2 e^6}{24\pi^4 \epsilon_0^3 c^3 m_e^2} \frac{1}{b^2 v^2}. \quad (39)$$

We could have guessed that the low frequency spectrum of the emission would be flat because, so far as these frequencies are concerned, the momentum impulse is a delta-function. It is a standard result of Fourier analysis that the

Fourier transform of the delta-function is a flat spectrum, $I(\omega) = \text{constant}$. To an excellent approximation, the low frequency spectrum is flat up to frequency $\omega = \gamma v/b$, above which the spectrum falls off exponentially.

Finally, we integrate over all relevant collision parameters which contribute to the radiation at frequency ω . Detailed calculations give the following answers for the spectral emissivity of the plasma at temperature T

$$\kappa_\nu = \frac{1}{3\pi^2} \left(\frac{\pi}{6}\right)^{1/2} \frac{Z^2 e^6}{\varepsilon_0^3 c^3 m_e^2} \left(\frac{m_e}{kT}\right)^{1/2} g(\nu, T) N N_e \exp\left(-\frac{h\nu}{kT}\right) \quad (40)$$

$$= 6.8 \times 10^{-51} Z^2 T^{-1/2} N N_e g(\nu, T) \exp(-h\nu/kT) \text{ W m}^{-3} \text{ Hz}^{-1} \quad (41)$$

where the number densities of electrons N_e and of nuclei N are given in m^{-3} . Suitable forms for the Gaunt factor at radio and X-ray wavelengths are:

$$\text{Radio } g(\nu, T) = \frac{\sqrt{3}}{2\pi} \left[\ln\left(\frac{128\varepsilon_0^2 k^3 T^3}{m_e e^4 \nu^2 Z^2}\right) - \gamma^{1/2} \right], \quad (42)$$

$$\text{X-ray } g(\nu, T) = \frac{\sqrt{3}}{\pi} \ln\left(\frac{kT}{h\nu}\right), \quad (43)$$

where $\gamma = 0.577\dots$ is Euler's constant. The total bremsstrahlung loss rate of the plasma is

$$-\left(\frac{dE}{dt}\right)_{\text{brems}} = 1.435 \times 10^{-40} Z^2 T^{1/2} \bar{g} N N_e \text{ W m}^{-3}. \quad (44)$$

where $\bar{g} \approx 1$. A compilation of a large number of useful Gaunt factors for a wide range of physical conditions is given by Karzas and Latter (1961). The evidence is wholly convincing that thermal bremsstrahlung is the emission mechanism responsible for the diffuse X-ray emission of clusters of galaxies and supernova remnants, clinching evidence being the detection of lines of FeXXVI in their X-ray spectra.

The development of similar techniques for treating relativistic and non-thermal bremsstrahlung is discussed in *HEA1*.

9.3 Cooling Flows in Clusters of Galaxies

In the central regions of clusters of galaxies, the gas density may become sufficiently high for the gas to cool by bremsstrahlung over cosmological time-scales. The observation of peaks in the X-ray surface brightness distribution in a number of clusters and the fact that the X-ray temperature is often lower in the

centre than in the outer regions are convincing pieces of evidence for presence of *cooling flows*. At the high temperatures present in the intracluster gas, the principal cooling mechanism is bremsstrahlung, for which the total energy loss rate is given by (44). Therefore, the cooling time-scale τ_{cool} is

$$\tau_{\text{cool}} = \frac{E}{\left| \frac{dE}{dt} \right|} = \frac{3kT}{1.435 \times 10^{-40} Z^2 T^{1/2} \bar{g} N^2} = \frac{10^{10} T_8^{1/2}}{N_4} \text{ years} \quad (45)$$

where T_8 is the temperature in units of 10^8 K, N_4 is the number density of electrons in units of 10^4 m^{-3} , $Z = 1$ and $\bar{g} = 1$. For the typical densities and temperatures found in the cores of clusters of galaxies, $T_8 \sim 1$, $N_e \geq 1$, the cooling time is less than the cosmological time-scale.

The observation of cooling flows is very important for many different aspects of the evolution of hot gas in clusters. The process adds mass to the central galaxy and may well be responsible for the formation of cool gas clouds in the central regions. In turn, this gas may be responsible for fuelling the massive black holes present in the most massive galaxies in clusters.

9.4 The Sunyaev–Zeldovich Effect in Hot Intracluster Gas

A completely independent method of studying the hot gas in clusters of galaxies is to search for decrements in the intensity of the Cosmic Microwave Background Radiation in the directions of clusters of galaxies. As the photons of the background radiation pass through the cloud, a few of them suffer Compton scattering by the hot electrons. Although, to first order, the photons are just as likely to gain as lose energy in these Compton scatterings, to second order there is a net statistical gain of energy. Thus, the spectrum of the Cosmic Microwave Background Radiation is shifted to slightly higher energies and, in the Rayleigh–Jeans region of the spectrum, there is expected to be a decrement in the intensity of the background radiation in the direction of the cluster of galaxies, while in the Wien region there should be a slight excess (Sunyaev and Zeldovich 1970). The magnitude of the distortion is determined by the *Compton scattering optical depth* through the region of hot gas

$$y = \int \left(\frac{kT_e}{m_e c^2} \right) \sigma_T N_e dl. \quad (46)$$

The resulting decrement in the Rayleigh–Jeans region of the spectrum is $\Delta I_\nu/I_\nu = -2y$ (see *HEA1*). Thus, the magnitude of the decrement along any line of sight through the cluster provides a measure of $\int N_e T_e dl$, in other words, the integral of the pressure of the gas along the line of sight. For the typical parameters of hot intracluster gas, the predicted decrement amounts to $\Delta I/I \approx 10^{-4}$.

The most impressive maps of the decrements in the Cosmic Microwave Background Radiation have been obtained by the Ryle Telescope at Cambridge. The contours defining the Sunyaev-Zeldovich decrements follow closely the X-ray brightness distribution of the hot cluster gas. In the case of Abell 1914, there is a region of high X-ray surface brightness close to the centre of the cluster and this is interpreted as a cooling flow. In conjunction with the X-ray bremsstrahlung measurements, these observations enable the physical conditions in the intracluster gas to be over-determined and so the physical dimensions of the hot gas clouds can be estimated, without knowledge of the distance of the cluster. This has enabled direct estimates of Hubble’s constant to be made.

10 Synchrotron Radiation

Synchrotron radiation, the emission of very high energy electrons gyrating in a magnetic field, is the process which dominates high energy astrophysics. The word *non-thermal* is used frequently in high energy astrophysics to describe the emission of high energy particles. This is unfortunate terminology since all emission mechanisms are ‘thermal’ in some sense. The word is conventionally taken to mean ‘continuum radiation from particles, the energy spectrum of which is not Maxwellian’. In practice, continuum emission is often referred to as ‘non-thermal’ if it cannot be described by the spectrum of thermal bremsstrahlung or black-body radiation.

10.1 Motion of a Relativistic Electron in a Uniform, Static Magnetic field

The motion of the particle consists of constant velocity along the magnetic field direction and circular motion with radius r about it. Thus, the particle moves in a *spiral path* with *constant pitch angle* θ . The radius r is known as the *gyroradius* of the particle. The angular frequency of the particle in its orbit ω_g is known as the *angular cyclotron frequency* or *angular gyrofrequency* and is given

by $\omega_g = v_\perp/r = zeB/\gamma m_0$ where $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor of the electron. The corresponding *gyrofrequency* ν_g is $\nu_g = \omega_g/2\pi = zeB/2\pi\gamma m_0$. In the case of a non-relativistic particle, $\gamma = 1$ and hence $\nu_g = zeB/2\pi m_0$. A useful figure to remember is the non-relativistic gyrofrequency of an electron $\nu_g = eB/2\pi m_e = 28 \text{ GHz T}^{-1}$ where the magnetic field strength is measured in tesla. Alternatively, $\nu_g = 2.8 \text{ MHz G}^{-1}$ for those not yet converted from gauss (G) to teslas (T).

10.2 The total energy loss rate

In synchrotron radiation, the acceleration is always perpendicular to the velocity vector of the particle and to \mathbf{B} . Therefore, the total radiation loss rate of the electron is

$$-\left(\frac{dE}{dt}\right) = \frac{\gamma^4 e^2}{6\pi\epsilon_0 c^3} |a_\perp|^2 = \frac{e^4 B^2}{6\pi\epsilon_0 c m_e^2} \frac{v^2}{c^2} \gamma^2 \sin^2 \theta \quad (47)$$

This can be rewritten in terms of the Thomson cross-section as

$$-\left(\frac{dE}{dt}\right) = 2\sigma_T c U_{\text{mag}} \left(\frac{v}{c}\right)^2 \gamma^2 \sin^2 \theta \quad (48)$$

where $U_{\text{mag}} = B^2/2\mu_0$ is the energy density of the magnetic field. In the ultrarelativistic limit, $v \rightarrow c$, we may safely set $v/c = 1$ in (48).

These results apply for electrons of a specific pitch angle θ . Particles of a particular energy E , or Lorentz factor γ , are often expected to have an isotropic distribution of pitch angles and therefore, averaging over an isotropic distribution of pitch angles $p(\theta) d\theta = \frac{1}{2} \sin \theta d\theta$

$$-\left(\frac{dE}{dt}\right) = \frac{4}{3}\sigma_T c U_{\text{mag}} \left(\frac{v}{c}\right)^2 \gamma^2 \quad (49)$$

There is a deeper sense in which (49) is the average loss rate for a particle of energy E . During its lifetime, it is likely that the high energy particle is randomly scattered in pitch angle and then (49) is the correct expression for its average energy loss rate.

10.3 The spectral distribution of radiation from a single electron - physical arguments

The next step is to work out the spectral distribution of synchrotron radiation. Let us analyse first of all some basic aspects of radiation mechanisms

of relativistic particles. One of the basic features of the radiation of relativistic particles is that the radiation is *beamed* in the direction of motion of the particle. Consider the case of a particle gyrating about the magnetic field at a pitch angle of 90° . The electron is accelerated radially inwards, and in the instantaneous rest frame emits the usual dipole pattern with respect to the acceleration vector. We can therefore work out the radiation pattern in the laboratory frame of reference using the relativistic aberration formulae. The angular distribution of radiation with respect to the velocity vector in the frame S' is $I_\nu \propto \sin^2 \theta' = \cos^2 \phi'$, where ϕ' is the angle between the velocity vector of the electron and the line of sight to the observer. The aberration formulae are:

$$\sin \phi = \frac{1}{\gamma} \frac{\sin \phi'}{1 + (v/c) \cos \phi'} \quad ; \quad \cos \phi = \frac{\cos \phi' + v/c}{1 + (v/c) \cos \phi'} \quad (50)$$

To illustrate the beaming of the radiation, consider the angles $\phi' = \pm\pi/4$, the angles at which the intensity of radiation falls to half its maximum value in the instantaneous rest frame of the particle. The corresponding angles ϕ in the laboratory frame of reference are $\sin \phi \approx \phi \approx 1/\gamma$.

Thus, the radiation emitted within $-\pi/4 < \phi' < \pi/4$ is beamed in the direction of motion of the electron within an angle $-1/\gamma < \phi < 1/\gamma$. As observed in the frame S , the dipole beam pattern is very strongly distorted and the intensity of the radiation is strongly Doppler-shifted. A 'spike' of radiation is observed every time the electron's velocity vector lies within an angle $\phi \sim 1/\gamma$ of the line of sight to the observer. The spectrum of the radiation is the Fourier transform of this pulse once the effects of time retardation are taken into account. The observer sees significant radiation from only about $1/\gamma$ of the particle's orbit, but the observed duration of the pulse is less than $1/\gamma$ times the period of the orbit because radiation emitted at the trailing edge of the pulse almost catches up with the radiation emitted at the leading edge.

Let us illustrate this effect by a simple calculation. Consider the observer located at a distance R from the position at which the leading edge of the pulse is first observed. The radiation reaches the observer at time R/c . Now consider the radiation emitted from the trailing edge of the pulse at time L/v later which then travels a distance $(R - L)$ at the speed of light to reach the observer. The trailing edge of the pulse arrives at the observer at a time $L/v + (R - L)/c$.

The duration of the pulse as measured by the observer is therefore

$$\Delta t = \left[\frac{L}{v} + \frac{(R-L)}{c} \right] - \frac{R}{c} = \frac{L}{v} \left(1 - \frac{v}{c} \right) \quad (51)$$

Thus, the observed duration of the pulse is much less than value L/v which might naively have been expected. We can rewrite (51) noting that

$$\frac{L}{v} = \frac{r_g \theta}{v} \approx \frac{1}{\gamma \omega_r} = \frac{1}{\omega_g} \quad (52)$$

where ω_g is the non-relativistic angular gyrofrequency and $\omega_r = \omega_g/\gamma$. We also note that we can rewrite $(1 - v/c)$ as

$$\left(1 - \frac{v}{c} \right) = \frac{[1 - (v/c)][1 + (v/c)]}{[1 + (v/c)]} = \frac{(1 - v^2/c^2)}{1 + (v/c)} \approx \frac{1}{2\gamma^2} \quad (53)$$

since $v \approx c$. Therefore, the observed duration of the pulse is roughly $\Delta t \approx 1/2\gamma^2\omega_g$. This means that the duration of the pulse as observed by a distant observer in the laboratory frame of reference is roughly $1/\gamma^2$ times shorter than the non-relativistic gyroperiod $T_g = 2\pi/\omega_g$. The maximum Fourier component of the spectral decomposition of the observed pulse of radiation corresponds to a frequency $\nu \sim \Delta t^{-1}$, that is, $\nu \sim \Delta t^{-1} \sim \gamma^2\nu_g$ where ν_g is the non-relativistic gyrofrequency. The same calculation can be performed for any pitch angle and then the result becomes $\nu \sim \gamma^2\nu_g \sin \theta$.

The reason for performing this simple exercise in detail is that the beaming of the radiation of ultrarelativistic particles is a very general property and does not depend upon the nature of the force causing the acceleration. Returning to an earlier part of the calculation, the observed frequency of the radiation can also be written

$$\nu \approx \gamma^2\nu_g = \gamma^3\nu_r = \frac{\gamma^3 v}{2\pi r_g} \quad (54)$$

where r_g is the radius of curvature of the particle's orbit. We may interpret r_g as the instantaneous radius of curvature of the particle's orbit. Thus, we can work out the frequency at which most of the radiation is emitted, provided we know the radius of curvature of the particle's orbit. The frequency of the observed radiation is roughly γ^3 times the angular frequency v/r where r is the instantaneous radius of curvature of the particle in its orbit. This result is

important in the study of *curvature radiation* which has important applications in the emission of radiation from the magnetic poles of pulsars.

For many calculations it is sufficient to know that the energy loss rate of the relativistic electron is given by (49) and that most of the radiation is emitted at a frequency $\nu \sim \gamma^2 \nu_g$ where ν_g is the non-relativistic gyrofrequency. The result of detailed calculations is that the emitted spectrum of a single electron, averaged over the particle's orbit is

$$j(\omega) = j_{\perp}(\omega) + j_{\parallel}(\omega) = \frac{\sqrt{3}e^3 B \sin \theta}{8\pi^2 \epsilon_0 c m_e} F(x) \quad (55)$$

where

$$x = \omega/\omega_c, \quad \omega_c = \frac{3}{2} \left(\frac{c}{v}\right) \gamma^3 \omega_r \sin \theta \quad \text{and} \quad F(x) = x \int_x^{\infty} K_{5/3}(z) dz. \quad (56)$$

ω_c is known as the critical angular frequency. $K_{5/3}(z)$ is a modified Bessel function of order 5/3. The spectrum has a broad maximum centred roughly at the frequency $\nu \approx \nu_c$ with $\Delta\nu/\nu \sim 1$. The maximum of the emission spectrum occurs at $\nu_{\max} = 0.29\nu_c$. The high frequency emissivity of the electron is given by an expression of the form $j(\nu) \propto \nu^{1/2} e^{-\nu/\nu_c}$, which is dominated by the exponential cut-off at frequencies $\nu \gg \nu_c$. The ratio of the powers emitted in the polarisations parallel and perpendicular to the magnetic field direction is $I_{\perp}/I_{\parallel} = 7$. To find the emissivity and polarisation observed from a distribution of electrons, we need to integrate over the energy spectrum of the emitting electrons.

10.4 The synchrotron radiation of a power law distribution of electron energies

The emitted spectrum of electrons of energy E is quite sharply peaked near the critical frequency ν_c and is much narrower than the breadth of the electron energy spectrum. Therefore, to a good approximation, it may be assumed that the radiation of an electron of energy E is radiated at the critical frequency ν_c

$$\nu \approx \nu_c \approx \gamma^2 \nu_g, \quad \nu_g = eB/2\pi m_e. \quad (57)$$

Therefore, the energy radiated in the frequency range ν to $\nu + d\nu$ can be attributed to electrons with energies in the range E to $E + dE$, which are assumed to have a power-law spectrum $N(E) = \kappa E^{-p}$. Therefore, the emissivity

is

$$J(\nu) d\nu = \left(-\frac{dE}{dt}\right) N(E) dE. \quad (58)$$

Now

$$E = \gamma m_e c^2 = \left(\frac{\nu}{\nu_g}\right)^{1/2} m_e c^2, \quad dE = \frac{m_e c^2}{2\nu_g^{1/2}} \nu^{-1/2} d\nu. \quad (59)$$

Substituting into (58) and using (49), the emissivity may be expressed in terms of κ , B , ν and fundamental constants.

$$J(\nu) = (\text{constants}) \kappa B^{(p+1)/2} \nu^{-(p-1)/2}. \quad (60)$$

The important result is that, if the electron energy spectrum has power law index p , the spectral index of the synchrotron emission, defined by $J(\nu) \propto \nu^{-\alpha}$, is $\alpha = (p - 1)/2$. The spectral shape is determined by the electron energy spectrum rather than by the shape of the emission spectrum of a single particle. The quadratic relation between emitted frequency and the energy of the electron accounts for the difference in slopes of the emission spectrum and the electron energy spectrum. The polarisation of the radiation from the power-law distribution of electrons is

$$\Pi = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{p + 1}{p + \frac{7}{3}}. \quad (61)$$

For the typical non-thermal source, for which $p \approx 2.6$, $\Pi = 73\%$.

10.5 Why is Synchrotron Radiation Taken so Seriously?

Synchrotron radiation dominates a great deal of thinking in high energy astrophysics and it is important to assess how convincing the evidence is.

- Perhaps the most important evidence comes from the comparison of the local flux of relativistic electrons measured at the top of the atmosphere with the predicted synchrotron radiation intensity if that flux of particles were present throughout the interstellar medium. This is not a trivial calculation since the electron spectrum in the relevant range of electron energies is strongly influenced by the effects of solar modulation. Nonetheless, the electron spectrum inferred from the Galactic radio emission can be fitted smoothly on to the region of the electron spectrum in which solar modulation can be neglected.

- A convincing case can be made that the high energy electrons are accelerated in supernova remnants. These are observed to be very strong radio sources with power-law intensity spectra and the radio emission is linearly polarised, similar to those of the diffuse radiation of the interstellar medium. Combining the frequency of occurrence of supernovae in our Galaxy with the typical energies they release in high energy particles, it is quite feasible to account for the energy requirements of the Galactic radio emission.
- The intense extragalactic radio sources have qualitatively similar power-law radio spectra and polarised radiation, but with intrinsic luminosities which are up to 10^8 greater than that of the Galaxy. The radio emission originates from enormous radio lobes rather than from the galaxy itself. The only reasonable way of accounting for the extended diffuse radio emission is that it is the synchrotron radiation of high energy electrons gyrating in magnetic fields within the radio lobes and that the particles were accelerated in the interaction of the beams of high energy particles from the nucleus with the ambient intergalactic medium.
- Direct evidence for relativistic particles in active galactic nuclei is provided by the very high brightness temperatures observed in compact radio sources which display synchrotron self-absorption.

10.6 Synchrotron Self-absorption

To every emission process there is a corresponding absorption process, in this case the process known as *synchrotron self-absorption*. Suppose a source of synchrotron radiation has a power law spectrum, $S_\nu \propto \nu^{-\alpha}$. Its *brightness temperature* is defined to be $T_b = (\lambda^2/2k)(S_\nu/\Omega)$, and is proportional to $\nu^{-(2+\alpha)}$, where S_ν is its flux density and Ω is the solid angle it subtends at the observer. The brightness temperature is the temperature of a black-body which would produce the observed surface brightness of the source at the frequency ν in the Rayleigh-Jeans limit, $h\nu \ll kT_e$. Thus, at low enough frequencies, the brightness temperature of the source approaches the kinetic temperature of the radiating electrons. When this occurs, self-absorption is important since thermodynamically the source cannot emit radiation of brightness temperature greater than its kinetic temperature.

The energy spectrum of the relativistic electrons is *not* a thermal equilibrium spectrum, but the concept of temperature can still be used. The characteristic time-scale for the relativistic electron gas to relax to an equilibrium spectrum is very long indeed under typical cosmic conditions because the particle number densities are very small and all interaction times with matter are very long. We can associate a temperature T_e with electrons of a given energy through the relativistic formula which relates particle energy to temperature $\gamma m_e c^2 = 3kT_e$. Thus, the *effective temperature* of the particles is a function of their energies. Since $\gamma \approx (\nu/\nu_g)^{1/2}$, $T_e \approx (m_e c^2/3k)(\nu/\nu_g)^{1/2}$.

For a self-absorbed source, the brightness temperature of the radiation is equal to the kinetic temperature of the emitting particles, $T_b = T_e$, and therefore, in the Rayleigh-Jeans limit,

$$S_\nu = \frac{2kT_e}{\lambda^2} \Omega = \frac{2m_e}{3\nu_g^{1/2}} \Omega \nu^{5/2} \propto \frac{\theta^2 \nu^{5/2}}{B^{1/2}}, \quad (62)$$

where Ω is the solid angle subtended by the source, $\Omega \approx \theta^2$. Thus, the low frequency spectrum of sources in which synchrotron self-absorption is important is $S_\nu \propto \nu^{5/2}$. It is a straightforward, but long, calculation to work out the absorption coefficient $\chi(\nu)$ for synchrotron self-absorption (see *HEA2*).

Many of the most compact radio sources have flat or inverted at centimetre wavelengths. VLBI observations show that the angular sizes of many of the synchrotron self-absorbed sources have angular sizes $\theta \approx 10^{-3}$ arcsec. For 1 Jy radio sources ($1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$), the corresponding brightness temperatures is

$$T_b = \frac{\lambda^2}{2k_B} \frac{S_\nu}{\Omega} \sim 10^{11} \text{ K}. \quad (63)$$

This is direct evidence for the presence of relativistic electrons within the source regions.

10.7 Distortions of Injection Spectra of the Electrons

In the optically thin regime of sources of synchrotron radiation, spectral breaks or cut-offs are often observed. In addition, different regions within individual sources may display spectral index variations. Both phenomena can be attributed to the effects of ageing of the spectrum of the electrons within the

source regions. An estimate of the lifetimes τ of the electrons in the source regions is

$$\tau = \frac{E}{(dE/dt)} = \frac{m_e c^2}{\frac{4}{3} \sigma_T c U_{\text{mag}} \gamma}. \quad (64)$$

For typical extended powerful radio sources, $\gamma \sim 10^3$ and $B \sim 10^{-9}$ T and so the lifetimes of the electrons are expected to be $\tau \leq 10^7 - 10^8$ years. In the case of X-ray sources, for example, the diffuse X-ray emission from the Crab Nebula and the jet of M87, the energies of the electrons are very much greater, the inferred magnetic field strengths are greater and so the relativistic electrons have shorter lifetimes. In these cases, the lifetimes of the electrons are significantly shorter than the light travel time across the sources and so the electrons must be accelerated within the source regions.

To obtain a quantitative description of the distortions of synchrotron radiation spectra, it is convenient to introduce the *diffusion-loss equation* for the electrons. If we write the loss rate of the electrons as $-(dE/dt) = b(E)$, the diffusion-loss equation is

$$\frac{\partial N(E)}{\partial t} = D \nabla^2 N(E) + \frac{\partial}{\partial E} [b(E) N(E)] + Q(E, t), \quad (65)$$

where D is a scalar diffusion coefficient and $Q(E)$ is a source term which describes the rate of injection of electrons into the source region (see *HEA2*).

Suppose there is an infinite, uniform distribution of sources, which inject high energy electrons with spectrum $Q(E) = \kappa E^{-p}$. Diffusion can be neglected and so, in the steady state,

$$N(E) = \frac{\kappa E^{-(p-1)}}{(p-1)b(E)}. \quad (66)$$

For high energy electrons under interstellar conditions, we can write

$$b(E) = A_1 \left(\ln \frac{E}{m_e c^2} + 19.8 \right) + A_2 E + A_3 E^2. \quad (67)$$

The first term on the right-hand side describes ionisation losses and depends only weakly upon energy; the second term represents bremsstrahlung and adiabatic losses and the last term describes inverse Compton and synchrotron losses (see *HEA2*). Thus, from (66),

- if ionisation losses dominate, $N(E) \propto E^{-(p-1)}$ – the energy spectrum is flatter by one power of E ;
- if bremsstrahlung or adiabatic losses dominate, $N(E) \propto E^{-p}$ – the spectrum is unchanged;
- if inverse Compton or synchrotron losses dominate, $N(E) \propto E^{-(p+1)}$ – the spectrum is steeper by one power of E .

These are also the equilibrium spectra expected whenever the continuous injection of electrons takes place over a time-scale longer than the lifetimes of the individual electrons involved.

These results find numerous applications in the study of non-thermal sources. In extended radio sources, the ‘hot-spots’ found towards the advancing edges of the extended source components often have flatter spectra than the extended radio lobes. Since the electrons are accelerated in the hot-spots, they are likely to be younger in these regions than in the extended lobes and the steepening of the spectra can be attributed to synchrotron losses in the extended source regions, enabling time-scales to be determined for the radio source. It is found that the γ -ray bursts have afterglow spectra which can be explained as the synchrotron emission of a relativistically expanding blast wave, exhibiting the effects of synchrotron losses and absorption at low frequencies (see Kulkarni 2002).

10.8 The Energetics of Sources of Synchrotron Radiation

An important calculation is the estimation of the minimum energy requirements in relativistic electrons and magnetic fields needed to account for the observed synchrotron emission. Suppose a source has luminosity $L_\nu \propto \nu^{-\alpha}$ and its volume is V . The luminosity can be related to the energy spectrum of the electrons and the magnetic field B in the source through the expression (60). Writing the energy density in relativistic electrons as ε_e , the total energy present in the source is

$$W_{\text{total}} = V\varepsilon_e + V\frac{B^2}{2\mu_0} = \int \kappa EN(E) dE + V\frac{B^2}{2\mu_0}. \quad (68)$$

The luminosity of the source L_ν determines only the product $V\kappa B^{1+\alpha}$. If V is known, the luminosity may either be produced by a large flux of relativistic

electrons in a weak magnetic field, or *vice versa*. Between the extremes, there is a minimum total energy.

In addition to relativistic electrons, there might be relativistic protons present as well. There are very few sources for which estimates of both the electron and proton fluxes are known. In our own Galaxy, there is about 100 times as much energy in relativistic protons as there is in electrons, whereas in the Crab Nebula, the energy in relativistic protons cannot be much greater than the energy in the electrons from dynamical arguments. To take account of the protons, it is customary to assume that they have energy β times that of the electrons, that is,

$$\varepsilon_{\text{protons}} = \beta\varepsilon_e, \quad \varepsilon_{\text{total}} = (1 + \beta)\varepsilon_e = \eta\varepsilon_e. \quad (69)$$

Therefore,

$$W_{\text{total}} = \eta V \int_{E_{\text{min}}}^{E_{\text{max}}} \kappa EN(E) dE + V \frac{B^2}{2\mu_0}. \quad (70)$$

Preserving only the essential dependences, it is straightforward to show that

$$W_{\text{total}} = G(\alpha)\eta L_\nu B^{-3/2} + V \frac{B^2}{2\mu_0}. \quad (71)$$

The minimum total energy is found by minimising (28) with respect to B .

$$B_{\text{min}} = \left[\frac{3\mu_0}{2} \frac{G(\alpha)\eta L_\nu}{V} \right]^{2/7}. \quad (72)$$

This magnetic field strength B_{min} corresponds to approximate equality of the energies in the relativistic particles and magnetic field. This condition is often referred to as *equipartition*. The minimum total energy is

$$W_{\text{total}}(\text{min}) = \frac{7}{6\mu_0} V^{3/7} \left[\frac{3\mu_0}{2} G(\alpha)\eta L_\nu \right]^{4/7}. \quad (73)$$

These estimates of B_{min} and W_{tot} are frequently used in the study of synchrotron sources of radio, optical and X-ray emission, but their limitations should be appreciated.

1. There is no physical justification for the source components being close to equipartition. It might be that the particle and magnetic field energies in the source components tend towards equipartition but there is no proof that this must be so.

2. The amount of energy present in the source is sensitive to the value of η , the amount of energy present in the form of relativistic protons.
3. The total energy in relativistic particles is dependent upon the limits assumed to the energy spectrum of the particles. There might be large fluxes of low energy electrons present with a quite different energy spectrum and we would have no way of knowing that they are present.
4. It has been assumed that the particles and magnetic field fill the source volume uniformly. The emitting regions might occupy only a small fraction of the apparent volume of the source, for example, if the emission originated in filaments or subcomponents within the volume V . Then, the volume would be smaller than V by a *filling factor* $f \ll 1$.
5. On the other hand, we can obtain a firm lower limit to the *energy density* within the source components since

$$U_{\min} = \frac{W_{\text{total}(\min)}}{V} = \frac{7}{6\mu_0} V^{-4/7} \left[\frac{3\mu_0}{2} G(\alpha)\eta L\nu \right]^{4/7}. \quad (74)$$

For dynamical purposes, the energy density is more important than the total energy since it is directly related to the relativistic gas pressure within the source components $p = \frac{1}{3}U$.

These estimates of the magnetic field strength and minimum energy should therefore be considered only order of magnitude estimates.

11 Inverse Compton scattering

Comptonisation is a vast subject (Pozdnyakov, Sobol and Sunyaev 1983). Inverse Compton scattering involves the scattering of low energy photons to high energies by ultrarelativistic electrons so that the photons gain and the electrons lose energy, the opposite of the standard Compton effect. We will treat the case in which the energy of the photon in the centre of momentum frame of the interaction is much less than $m_e c^2$, and consequently the Thomson cross-section can be used to describe the scattering probability.

Many of the most important results can be worked out using simple physical arguments (see, for example, Blumenthal and Gould 1970 and Rybicki and Lightman 1979). Consider the collision between a photon and a relativistic

electron as seen in the laboratory frame of reference S and in the rest frame of the electron S'. Since $\gamma\hbar\omega \ll m_e c^2$, the centre of momentum frame is essentially that of the electron. If the energy of the photon is $\hbar\omega$ and the angle of incidence θ in S, its energy in the frame S' is

$$\hbar\omega' = \gamma\hbar\omega[1 + (v/c)\cos\theta] \quad (75)$$

according to the standard Doppler shift formula. Similarly, the angle of incidence θ' in the frame S' is related to θ by the aberration formulae

$$\sin\theta' = \frac{\sin\theta}{\gamma[1 + (v/c)\cos\theta]} \quad ; \quad \cos\theta' = \frac{\cos\theta + v/c}{1 + (v/c)\cos\theta}. \quad (76)$$

In the rest frame of the electron, the interaction is simply Thomson scattering and hence the energy loss rate of the electron in S' is the rate at which energy is reradiated by the electron. From the considerations of Sect. 8.1, the energy loss rate is

$$-(dE/dt)' = \sigma_T c U'_{\text{rad}}, \quad (77)$$

where U'_{rad} is the energy density of radiation in the rest frame of the electron. It is left as an exercise to the reader to show that

$$U'_{\text{rad}} = \frac{4}{3} U_{\text{rad}} (\gamma^2 - \frac{1}{4}). \quad (78)$$

Because $(dE/dt) = (dE/dt)'$, we find

$$dE/dt = \frac{4}{3} \sigma_T c U_{\text{rad}} (\gamma^2 - \frac{1}{4}). \quad (79)$$

This is the energy gained by the photon field due to the scattering of the low energy photons. The rate at which energy is removed from the low energy photon field is $\sigma_T c U_{\text{rad}}$ and therefore, subtracting, we find

$$dE/dt = \frac{4}{3} \sigma_T c U_{\text{rad}} (\gamma^2 - \frac{1}{4}) - \sigma_T c U_{\text{rad}} = \frac{4}{3} \sigma_T c U_{\text{rad}} (\gamma^2 - 1). \quad (80)$$

Using the identity $(\gamma^2 - 1) = (v^2/c^2)\gamma^2$, the loss rate in its final form is

$$dE/dt = \frac{4}{3} \sigma_T c U_{\text{rad}} \left(\frac{v^2}{c^2} \right) \gamma^2. \quad (81)$$

Notice the remarkable similarity between the expressions for the loss rates by synchrotron radiation (49) and by inverse Compton scattering (81). In both

cases, the electron is accelerated by the electric field which it observes in its instantaneous rest-frame. In the case of synchrotron radiation, the constant accelerating electric field is associated with the motion of the electron through the magnetic field, $\mathbf{E}' = \mathbf{v} \times \mathbf{B}$, and, in the case of inverse Compton scattering, it is the sum of all the electric fields of the incident waves.

The intensity spectrum for an incident isotropic photon field at a single frequency ν_0 is derived by Blumenthal and Gould (1970). At low frequencies, the scattered radiation has the form $I(\nu) \propto \nu$. The maximum energy which the photon can acquire corresponds to a head-on collision in which the photon is sent back along its original path,

$$(\hbar\omega)_{\max} = \hbar\omega\gamma^2(1 + v/c)^2 \approx 4\gamma^2\hbar\omega_0. \quad (82)$$

The average energy of the scattered photons is

$$\hbar\omega = \frac{4}{3}\gamma^2(v/c)^2\hbar\omega_0 \approx \frac{4}{3}\gamma^2\hbar\omega_0. \quad (83)$$

The general result that the frequency of the scattered photons is $\nu \approx \gamma^2\nu_0$ is of profound importance in high energy astrophysics. We know that there are electrons with Lorentz factors $\gamma \sim 100 - 1000$ in astronomical sources and consequently they scatter any low energy photons to very much higher energies. Consider radio, infrared and optical photons scattered by electrons with $\gamma = 1000$, so that the scattered radiation has frequency (or energy) 10^6 times that of the incoming photons. Thus, radio photons with $\nu_0 = 10^9$ Hz become ultraviolet photons with $\nu = 10^{15}$ Hz ($\lambda = 300$ nm); far-infrared photons with $\nu_0 = 3 \times 10^{12}$ Hz, produce X-rays with frequency 3×10^{18} Hz (about 10 keV); optical photons with $\nu_0 = 4 \times 10^{14}$ Hz become γ -rays with frequency 4×10^{20} Hz (about 1.6 MeV). It is apparent that the inverse Compton scattering process is a means of producing very high energy photons indeed.

When these formulae are used in astrophysical calculations, it is necessary to integrate over both the spectrum of the incident radiation and the spectrum of the relativistic electrons (see Blumenthal and Gould 1970). Some of the results are immediately apparent from the analogy between the inverse Compton scattering and synchrotron radiation processes. For example, the spectrum of the inverse Compton scattering of photons of energy $h\nu$ by a power-law distribution of electron energies $dN \propto E^{-p} dE$, by analogy with the results of the

calculation which resulted in (60), is

$$I(\nu) \propto \nu^{-(p-1)/2}, \quad (84)$$

because of the γ^2 dependence of the energy loss rate by inverse Compton scattering and the fact that the frequency of the scattered radiation is $\nu \approx \gamma^2 \nu_0$.

The ratio of the total amount of energy liberated by synchrotron radiation process and by inverse Compton scattering by the same distribution of electrons is

$$\frac{(dE/dt)_{\text{sync}}}{(dE/dt)_{\text{IC}}} = \frac{\int I_\nu d\nu \text{ (radio)}}{\int I_X d\nu_X \text{ (X-ray)}} = \frac{B^2/2\mu_0}{U_{\text{rad}}}, \quad (85)$$

where U_{rad} is the energy density of radiation and B the magnetic flux density in the source region. Thus, if we measure the radio and X-ray flux densities from a source region and we know U_{rad} , we can find the magnetic flux density in the source. Diffuse X-ray emission from the extended radio lobes has been searched for in the cases of the bright radio sources Cygnus A, Centaurus A and Fornax A due to inverse Compton scattering of photons of the Cosmic Microwave Background Radiation. It has proved difficult to find convincing evidence for this process from the same population of electrons responsible for the radio emission. Probably the most convincing case is that of Fornax A in which the X-ray emission is coincident with the radio lobes (Feigelson *et al.* 1996). The inferred magnetic field strength is $B \approx 2 - 3 \times 10^{-10}$ T, a value close to that derived from equipartition arguments.

Another important result involving the Cosmic Microwave Background Radiation is that relativistic electrons can never escape from it since it permeates all space. The energy density of the Cosmic Microwave Background Radiation is $U_0 = aT^4 = 2.6 \times 10^5$ eV m⁻³ and so the maximum lifetime τ of any electron is

$$\tau = \frac{E}{|dE/dt|} = \frac{E}{\frac{4}{3}\sigma_T c \gamma^2 U_0} = \frac{2.3 \times 10^{12}}{\gamma} \text{ years}. \quad (86)$$

100 GeV electrons are observed at the top of the atmosphere and so they must have lifetimes $\tau \leq 10^7$ years.

12 Synchro-Compton Radiation and the Inverse Compton Catastrophe

Wherever there are large number densities of soft photons, the presence of ultrarelativistic electrons results in the production of high energy photons. A case of special interest is that in which the same relativistic electrons which are the source of the soft photons are also responsible for scattering these photons to X-ray and γ -ray energies – this is the process known as *synchro-Compton Radiation*. One case of special importance is that in which the number density of low energy photons is so great that most of the energy of the electrons is lost by synchro-Compton radiation rather than by synchrotron radiation. This line of reasoning leads to what is known as the *inverse Compton catastrophe*.

The ratio, η , of the rates of loss of energy of an ultrarelativistic electron by inverse Compton and synchrotron radiation in the presence of a photon energy density U_{rad} and a magnetic field of magnetic flux density B is

$$\eta = \frac{(dE/dt)_{\text{IC}}}{(dE/dt)_{\text{sync}}} = \frac{U_{\text{photon}}}{B^2/2\mu_0}. \quad (87)$$

The synchro-Compton catastrophe occurs if this ratio is greater than 1. In this case, the energy density of the X-rays is greater than that of the radio photons and so the electrons suffer even greater energy losses by scattering these X-rays to γ -ray energies. In turn, these γ -rays have a greater energy density than the X-rays ... and so on. Thus, as soon as η becomes greater than one, all the energy of the electrons is lost at the very highest energies and the radio source should be a very powerful source of X-rays and γ -rays

For a source which is synchrotron self-absorbed, the ratio of the loss rates due to synchro-Compton radiation η can be shown to be

$$\eta = \frac{(dE/dt)_{\text{IC}}}{(dE/dt)_{\text{sync}}} = \left(\frac{81e^2\mu_0 k_{\text{B}}^5}{\pi^2 m_{\text{e}}^6 c^{11}} \right) \nu T_{\text{e}}^5 \quad (88)$$

(see *XSA*). Thus, the ratio of the loss rates depends very strongly upon the brightness temperature of the radio source. Putting in the values of the constants, the critical brightness temperature at which $\eta = 1$ is

$$T_{\text{b}} = T_{\text{e}} = 10^{12} \nu_9^{-1/5} \text{ K}, \quad (89)$$

where ν_9 is the frequency at which the brightness temperature is measured in GHz. According to this calculation, no compact radio source should have

brightness temperature greater than $T_B \approx 10^{12}$ K, if the emission is incoherent synchrotron radiation.

The most compact sources studied by VLBI at centimetre wavelengths have brightness temperatures which are less than the synchro-Compton limit, typically, the values found being $T_B \approx 10^{11}$ K. As noted above, this is direct evidence that the radiation is the emission of relativistic electrons since the temperature of the emitting electrons must be at least 10^{11} K. This is not, however, the whole story. If the time-scales of variability τ of the compact source is used to estimate its physical sizes, $l \sim c\tau$, the source regions must be considerably smaller than those inferred from VLBI, and values of T_B exceeding 10^{12} K are found. It is likely that relativistic beaming is the source of this discrepancy (see Sect. 14).

There is no definite evidence that synchro-Compton radiation has been observed in any of the X-ray and γ -ray sources, but it would certainly be no surprise if it were the origin of the emission in the intense γ -ray sources observed by the CGRO. There is evidence that radio quasars have greater X-ray luminosities than radio quiet quasars and synchro-Compton radiation may well be involved. Examples of the expected spectra of sources of synchro-Compton radiation have been evaluated by Band and Grindlay (1985) for both homogeneous and inhomogeneous sources. A number of important refinements are included in their computations, including the fact that, at relativistic energies $h\nu \geq 0.5$ MeV, the Klein-Nishina cross-section rather than the Thomson cross-section should be used for photon-electron scattering. In the ultrarelativistic limit, the cross-section is

$$\sigma_{\text{KN}} = \frac{\pi^2 r_e^2}{h\nu} \left(\ln 2h\nu + \frac{1}{2} \right), \quad (90)$$

and so the cross-section decreases as $(h\nu)^{-1}$ at high energies. Consequently, higher order scatterings result in much reduced luminosities as compared with the non-relativistic calculation.

13 γ -ray Processes, Photon-photon Interactions and the Compactness Parameter

Synchrotron radiation, inverse Compton scattering and relativistic bremsstrahlung are effective means of creating high-energy γ -ray photons, but there are other mechanisms.

13.1 Neutral pion decay

One of the most important is the decay of neutral pions created in collisions between relativistic protons and nuclei of atoms and ions of the interstellar gas.

$$p + p \rightarrow \pi^+, \pi^-, \pi^0. \quad (91)$$

The charged pions decay into muons and neutrinos

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \quad ; \quad \pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (92)$$

with a mean lifetime of 2.551×10^{-8} s. The charged muons then decay with mean lifetime of 2.2001×10^{-6} s

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \quad ; \quad \mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu. \quad (93)$$

while the neutral pions decay into pairs of γ -rays, $\pi^0 \rightarrow \gamma + \gamma$, in only 1.78×10^{-16} s. The cross-section for this process is $\sigma_{pp \rightarrow \gamma\gamma} \approx 10^{-30}$ m² and the emitted spectrum of γ -rays has a broad maximum centred on a γ -ray energy of about 70 MeV (see *HEA2*). This is the process responsible for the continuum emission of the interstellar medium at energies $\varepsilon \geq 100$ MeV.

13.2 Photon-photon collisions

It is a useful exercise in special relativity to show that the threshold energy for the collision of two photons with energies ε_1 and ε_2 to create an electron-positron pair is

$$\varepsilon_2 \varepsilon_1 = \frac{2m_e^2 c^4}{(1 - \cos \theta)}, \quad (94)$$

where θ is the angle between the incident directions of the photons. The threshold for this process occurs for head-on collisions, $\theta = \pi$ and hence,

$$\varepsilon_2 \geq \frac{m_e^2 c^4}{\varepsilon_1} = \frac{0.26 \times 10^{12}}{\varepsilon_1} \text{ eV}, \quad (95)$$

where ε_1 is measured in electron volts. This process provides not only a means for creating electron-positron pairs, but also results an important source of opacity for high-energy γ -rays, in particular, for sources in which there are

large fluxes of γ -rays with energy about 1 MeV. In the limit $\hbar\omega \approx m_e c^2$, the cross-section for this process is

$$\sigma = \pi r_e^2 \left(1 - \frac{m_e^2 c^4}{\omega^2}\right)^{1/2} \quad (96)$$

(Ramama Murthy and Wolfendale 1986). Near threshold, the cross-section for the interaction $\gamma\gamma \rightarrow e^+e^-$ is $\sigma \sim \pi r_e^2 \sim 0.2\sigma_T$. These cross-sections enable the opacity of the interstellar and intergalactic medium to be evaluated as well as providing a mechanism by which large fluxes of positrons could be generated in the vicinity of active galactic nuclei.

13.3 The compactness parameter

These considerations are particularly important in the case of the extremely luminous and highly variable extragalactic γ -ray sources discovered by the Compton Gamma-Ray Observatory (CGRO). A key role is played by the *compactness parameter*, which arises in considerations of whether or not a γ -ray source is opaque for $\gamma\gamma$ collisions because of pair production. We carry out the calculation for the flux of γ -rays at threshold, $\varepsilon \sim m_e c^2$. The mean free path of the γ -ray for $\gamma\gamma$ collisions is $\lambda = (N_\gamma \sigma)^{-1}$ where N_γ is the number density of photons with energies $\varepsilon = h\nu \sim m_e c^2$. If the source has luminosity L_γ and radius r , the number density of photons within the source region is $N_\gamma = L_\gamma / 4\pi r^2 c \varepsilon$. The condition that the source is opaque is $r \approx \lambda$, that is,

$$r \sim \frac{4\pi r^2 c m_e c^2}{L_\nu \sigma}, \quad \text{or} \quad \frac{L_\nu \sigma}{4\pi m_e c^3 r} \sim 1. \quad (97)$$

The compactness factor C is defined to be the quantity

$$C = \frac{L_\nu \sigma}{4\pi m_e c^3 r} \quad (98)$$

Sometimes the compactness parameter is defined without the factor of 4π in the denominator. If the compactness parameter is very much greater than unity, the γ -rays are all destroyed by electron-positron pair production, resulting in a huge flux of electrons and positrons within the source region. Consequently, the source would no longer be a hard γ -ray source. The importance of the compactness parameter can be appreciated from observations of some of the intense variable γ -ray sources observed by the CGRO. These have enormous

luminosities, $L_\gamma \sim 10^{41}$ W and have been observed to vary significantly in intensity over time-scales of a few days. Inserting these values into (98), $C \gg 1$ and so there is a problem in understanding why these sources exist. It turns out that all the ultraluminous γ -ray sources are associated with compact radio sources, which exhibit synchrotron self-absorption and many of these display superluminal motions. The inference is that the luminosities of the γ -ray sources and the time-scales of variation have been significantly changed by the effects of relativistic beaming.

14 Relativistic Beaming

We have encountered a number of cases in which relativistic beaming is important. These include:

- The observation of superluminal radio sources with component separation speeds up to $10c$. These include the microquasars in our Galaxy.
- The very rapid variability of BL Lac objects and blazars.
- The ultraluminous γ -ray sources and their variability.
- The avoidance of the inverse Compton catastrophe.

Let us begin with the most popular model for superluminal sources, the *relativistic ballistic model*. The aim is to determine the observed transverse speed of a component ejected at some angle θ to the line of sight at velocity v . The observer is located at a distance D from the source. The source component is ejected from the origin O at some time t_0 and the signal from that event arrives at the observer at time $t = D/c$ later. After time t_1 , the component is located at a distance vt_1 from the origin and so is observed at a projected distance $vt_1 \sin \theta$ according to the distant observer. The light signal bearing this information arrives at the observer at time

$$t_2 = t_1 + \frac{D - vt_1 \cos \theta}{c}, \quad (99)$$

since the signals have to travel a slightly shorter distance $D - vt_1 \cos \theta$ to reach the observer. Therefore, according to the distant observer, the transverse speed

of the component is

$$v_{\perp} = \frac{vt_1 \sin \theta}{t_2 - t} = \frac{v \sin \theta}{1 - \frac{v \cos \theta}{c}}. \quad (100)$$

The maximum apparent transverse speed occurs at $\cos \theta = v/c$ and is $v_{\perp} = \gamma v$, where $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor. Thus, provided the source component moves at a speed close enough to the speed of light, apparent motions on the sky $v_{\perp} \gg c$ can be observed without violating the postulates of special relativity. For example, if the source component were ejected at $0.98c$, transverse velocities up to $\gamma c = 5c$ are perfectly feasible.

The trickier part of the story is to understand the effects of what is loosely referred to as ‘relativistic beaming’ upon the observed intensities of the source components. Let us consider first an undergraduate problem in relativity:

- *A rocket travels towards the Sun at speed $v = 0.8c$. Work out the luminosity, colour, angular size and brightness of the Sun as observed from the spaceship when it crosses the orbit of the Earth. It may be assumed that the Sun radiates like a uniform disc with a black-body spectrum at temperature T_0 .*

This problem includes many of the effects found in relativistic beaming problems. Let us work out the separate effects involved in evaluating the intensity of radiation observed in the moving frame of reference. Consider the radiation from an annulus of angular width $\Delta\theta$ at angle θ with respect to the centre of the Sun. There are four changes to consider:

- *The frequency shift of the radiation $\nu' = \gamma\nu_0 [1 + (V/c) \cos \theta] = \kappa\nu_0$. This is the expression for the ‘blue-shift’ of the radiation due to the motion of the spacecraft.*
- *The waveband $\Delta\nu$, in which the radiation is observed, is blue shifted by the same factor $\Delta\nu' = \kappa\Delta\nu_0$.*
- *Time intervals are also different between the stationary and moving frames. Comparing the periods of the waves as observed in S and S', $\nu' = 1/T'$, $\nu_0 = 1/T_0$, and so $T'/T = \nu_0/\nu'$. Since the periods T and T' can be considered to be the times measured on clocks in their respective*

frames, the radiation emitted in the time interval Δt is observed in the time interval $\Delta t'$ by the observer in S' such that $\Delta t' = \Delta t/\kappa$.

- *Solid Angles* Finally, we work out how the solid angle subtended by the annulus changes between the two frames of reference.

$$\sin \theta' d\theta' d\phi' = \frac{\sin \theta d\theta d\phi}{\kappa^2} \quad d\Omega' = \frac{d\Omega}{\kappa^2}. \quad (101)$$

Thus, the solid angle in S' is smaller by a factor κ^2 as compared with that observed in S . Exactly the same form of beaming appears in the derivation of the formulae for synchrotron radiation.

We can now put these results together to work out how the intensity of radiation from the region of the Sun within solid angle $d\Omega$ changes between the two frames of reference. First of all, the intensity $I(\nu)$ is defined to be the power arriving at the observer per unit frequency interval per unit solid angle from the direction θ . The observer in the spacecraft observes the radiation arriving in the solid angle $d\Omega'$ about the angle θ' and we need to transform its other properties to those observed in S' . Let us enumerate how the factors change the observed intensity. The energy $h\nu N(\nu)$ received in S in the time interval Δt , in the frequency interval $\Delta\nu$ and in solid angle $\Delta\Omega$ is observed in S' as an energy $h\nu' N(\nu')$ in the time interval $\Delta t'$, in the frequency interval $\Delta\nu'$ and in solid angle $\Delta\Omega'$, where $N(\nu) = N(\nu')$ is the invariant number of photons. Therefore, the intensity observed in S' is

$$I(\nu') = I(\nu) \times \frac{\kappa \times \kappa \times \kappa^2}{\kappa} = I(\nu)\kappa^3. \quad (102)$$

Now, apply this result to the spectrum of black-body radiation,

$$I(\nu) = \frac{2h\nu^3}{c^2} \left(e^{h\nu/kT} - 1 \right)^{-1}. \quad (103)$$

Then,

$$I(\nu') = \frac{2h\nu'^3 \kappa^3}{c^2} \left(e^{h\nu/kT} - 1 \right)^{-1} = \frac{2h\nu'^3}{c^2} \left(e^{h\nu'/kT'} - 1 \right)^{-1}, \quad (104)$$

where $T' = \kappa T$. In other words, the observer in S' observes a black-body radiation spectrum with temperature $T' = \kappa T$. Inserting the values $v = 0.8c$,

$\theta = 0^\circ$, we find $\kappa = 3$ and so the moving observer finds the Sun to have three times greater temperature than the observer on Earth.

(104) also describes the temperature distribution of the Cosmic Microwave Background Radiation over the sky as observed by the COBE observatory. At the mK intensity level, it is found that its temperature distribution follows very precisely a dipole distribution, $T = T_0[1 + (V/c) \cos \theta]$. This is interpreted as an indication that the Solar System is moving through the frame of reference in which the Cosmic Microwave Background Radiation would be perfectly isotropic at a velocity of about 600 km s^{-1} . Since $V/c \approx 2 \times 10^{-3}$, $\gamma \approx 1$.

For relativistically moving source components, we need to determine the value of κ for a source component moving at velocity V at an angle θ with respect to the line of sight from the observer to the source. A straightforward calculation shows that

$$\kappa = \frac{1}{\gamma \left(1 - \frac{V \cos \theta}{c}\right)}, \quad (105)$$

if the source is moving towards the observer. As in the above example, the observed flux density of the source is

$$S(\nu_{\text{obs}}) = \frac{L(\nu_0)}{4\pi D^2} \times \kappa^3, \quad (106)$$

where $\nu_{\text{obs}} = \kappa \nu_0$. For non-thermal sources, the spectra can often be described by a power-law $L(\nu_0) \propto \nu_0^{-\alpha}$ and so

$$S(\nu_0) = \frac{L(\nu_0)}{4\pi D^2} \times \kappa^{3+\alpha}. \quad (107)$$

If the superluminal sources consisted of identical components ejected from the radio source at the same angle in opposite directions, the relative intensities of the two components would be in the ratio

$$\frac{S_1}{S_2} = \left[\frac{1 + (v/c) \cos \theta}{1 - (v/c) \cos \theta} \right]^{3+\alpha}. \quad (108)$$

There should therefore be large differences in the observed intensities of the jets. For example, if we adopt the largest observed velocities for a given value of γ , $\cos \theta = v/c$, then in the limit $v \approx c$,

$$\frac{S_1}{S_2} = (2\gamma^2)^{3+\alpha}. \quad (109)$$

Since values of $\gamma \sim 10$ are observed and $\alpha \sim 0-1$, it follows that the advancing component would be very much more luminous than the receding component. It is, therefore, not at all unexpected that the sources should be one-sided.

Another complication is the fact that the emission is often associated with jets. Care has to be taken because, if the jet as a whole is moving at velocity v , then the time dilation formula $\Delta t' = \Delta t/\kappa$ shows that the advancing component is observed in a different proper time interval as compared with the receding component, the time which has passed in the frame of the source being $\Delta t_1 = \Delta t_0/\kappa$ where Δt_0 is the time measured in the observer's frame of reference. If the jet consisted of a stream of components ejected at a constant rate from the active galactic nucleus, the observed intensity of the jet would be enhanced by a factor of only $\kappa^{2+\alpha}$. Thus, the precise form of the relativistic beaming factor is model dependent and care needs to be taken about the assumptions made.

Let us apply these considerations to the cases of sources exceeding the limiting surface brightness $T_b = 10^{12}$ K discussed in Sect. 12 and the compactness parameter discussed in Sect. 13.3. In the case of the Inverse Compton Catastrophe, (88) shows that the ratio of the loss rates for inverse Compton scattering and synchrotron radiation depends upon the product νT_b^5 . Since the brightness temperature $T_{\text{obs}} = \kappa^5 T_0$ and $\nu_{\text{obs}} = \kappa \nu_0$, it follows that $\eta \propto \kappa^6$ and so the observed value of T_b can exceed 10^{12} K if the source is moving at such a high velocity that $\kappa \gg 1$.

In the case of the compactness parameter,

$$C = \frac{L_\nu \sigma_T}{4\pi m_e c^3 \times ct}, \quad (110)$$

the relativistic beaming factors enable us to understand why these sources should exist. In (110), it is assumed that the dimensions of the source are $l \approx ct$ from their rapid time variability. The observed luminosity is enhanced by a factor $\kappa^{3+\alpha}$ and, in addition, because the time-scale of variability appears on the denominator of (110), the compactness parameter is increased by relativistic beaming by a factor of roughly $\kappa^{4+\alpha}$. Since $\alpha \approx 1$, $C \propto \kappa^5$ and so, in the frame of the source components themselves, the value of the compactness parameter can be reduced below the critical value.

References

1. Antonucci, R. (1993). *ARAA*, **31**, 473.
2. Band, D.L. and Grindlay, J.E. (1985). *ApJ*, **298**, 128.
3. Berezhko, E. (2002). Proc. International School Space Science (ISSS), *Astroparticle and Gamma ray Physics in Space*, (eds. A. Morselli and P. Piccozza), (this volume).
4. Best, P.N., Röttgering, H.J.A. and Longair, M.S. (2000). *MNRAS*, **311**, 1 and 23.
5. Blandford, R.D., Netzer, H. and Woltjer, L. (1990). Berlin: Springer-Verlag.
6. Blumenthal, G.R. and Gould, R.J. (1970). *Rev. Mod. Phys.*, **42**, 237.
7. Böhringer, H. (1995). *Ann. New York Acad. Sciences*, **759**, 67.
8. Feigelson, E.P., Laurent-Muehleisen, S.A., Kollgaard, R.I. and Fomalont, E.B. (1996). *ApJ*, **449**, 149.
9. Ghez, A.M., Morris, M., Becklin, E.E., Tanner, A. and Kremenek, T. (2000). *Nature*, **407**, 349.
10. Jackson, J.D. (1975). *Classical Electrodynamics*. New York: Wiley and Sons, Inc.
11. Karzas, W.J. and Latter, R. (1961). *ApJS*, **6**, 167.
12. Kellermann, K.I. (2001). In *Quasars, AGNs and Related Research Across 2000*, (eds. G. Setti and J.-P. Swings), p. 26. Berlin: Springer-Verlag.
13. Kulkarni, S. (2002). In *Lighthouses of the Universe*, (eds. R.A. Sunyaev and M. Gilfanov). Berlin: Springer-Verlag, (in press).
14. Longair, M.S. (1992) (revised edition 1997) *High Energy Astrophysics. Vol. 1.* Cambridge: Cambridge University Press (*HEA1*).
15. Longair, M.S. (1994) (revised edition 1997) *High Energy Astrophysics. Vol. 2.* Cambridge: Cambridge University Press (*HEA2*).

16. Longair, M.S. (1999). In *X-ray Spectroscopy in Astrophysics*, (eds. J. van Paradijs and J.A.M. Bleeker), p 1. Berlin: Springer-Verlag. (*XSA*)
17. Mirabel, I.F. and Rodriguez, L.F. (1998). *Nature*, **392**, 673.
18. Pozdnyakov, L.A., Sobel, I.M. and Sunyaev, R.A. (1983). *Astrophys. Sp. Phys. Reviews*, **2**, 263.
19. Ramana Murthy, P. and Wolfendale, A.W. (1986, 1993). *Gamma-ray Astronomy*. Cambridge: Cambridge University Press.
20. Rybicki, G.B. and Lightman, A.P. (1979). *Radiative Processes in Astrophysics*. New York: Interscience Publishers.
21. Sunyaev, R.A. and Zeldovich, Ya.B. (1970). *Astrophys. Sp. Phys.*, **7**, 20.

**ASTROPARTICLES AND COSMOLOGY:
LEARNING FROM COSMIC AND GAMMA RAYS**

John Ellis

Theoretical Physics Division, CERN, CH 1211 Geneva 23

ABSTRACT

After a brief introduction, three topics at the interface between particle physics, astrophysics, cosmology and cosmic-ray physics are discussed. First, the lightest supersymmetric particle is proposed as a plausible candidate for cold dark matter, and the prospects for detecting it among the cosmic rays are reviewed. Secondly, there could also be superheavy metastable particle relics from the Big Bang, whose decays might be responsible for ultrahigh-energy cosmic rays. Finally, the speculative possibility of using photons from gamma-ray bursters and other astrophysical sources to probe models of quantum gravity is mentioned.

1 Supersymmetric Dark Matter

1.1 Cosmic Probes of Physics Beyond the Standard Model

The main focus in these lectures is on signatures in cosmic and gamma rays for physics beyond the Standard Model, so I first review ideas what this physics might be, and recall how easily cosmology might provide windows on its nature. I remind you that all confirmed particle physics experiments at accelerators agree with the Standard Model, which has passed many tests at better than the percent level. However, the Standard Model is theoretically very unsatisfactory. We would like to unify the fundamental interactions in a single Grand Unified Theory (GUT). We would like to understand the proliferation of different types of quarks and leptons and the pattern of their weak interactions. We would like to understand the origin of particle masses, which is generally thought to be a Higgs boson, and why they are so much smaller than the Planck mass: this is a problem where supersymmetry may play a rôle ¹⁾. Finally, beyond all these questions, we would like to construct a Theory of Everything (TOE) including gravity, which should reconcile it with quantum mechanics: the prime candidate for the TOE is string theory, which also seems to require supersymmetry for its consistency.

Astrophysics and cosmology provide many arenas for probing the answers to these questions, some of which may lie beyond the reach of foreseeable accelerators. For example, the scale of *cosmological inflation* seems to be close to the GUT scale. GUTs and flavour theories may also be tested through their predictions for *neutrino masses and mixing*, which are potentially important for astrophysics (e.g., supernovae) and cosmology (e.g., structure formation). One of the prime candidates for *cold dark matter* is the lightest supersymmetric particle ²⁾. Another candidate is some metastable superheavy relic particle, whose decays might be responsible for the *ultrahigh-energy cosmic rays* ³⁾, and which might be a spin-off from some variant of superstring theory ⁴⁾. As a final example, modifications of special relativity, such as might appear in a quantum theory of gravity ⁵⁾, may best be tested using astrophysical sources such as *gamma-ray bursters* ⁶⁾.

High-energy astrophysical sources are described in parallel lectures, so

here I make a brief introduction to early cosmology, setting the other stage for probing fundamental physics.

During the normal adiabatic expansion of the Universe, the temperature $T \sim 1/a$, where a is the scale size of the Universe. Thus the early, smaller Universe was correspondingly hotter. Before decoupling, it is thought that most of the energy in the Universe was composed of relativistic particles, in which case the age $t \propto R^2 \propto 1/T^2$, with characteristic particle energies $E \sim T$. Putting in the units, the time-mass energy relation is approximately

$$t \text{ (seconds)} \simeq \frac{\mathcal{O}(1)}{T(\text{MeV})^2} \quad (1)$$

where the $\mathcal{O}(1)$ coefficient depends on the total number of particle species. Thus, when the Universe was about 1 second old, and Big-Bang nucleosynthesis started, the temperature was about 10^{10} K, corresponding to particle energies ~ 1 MeV, comparable to the electron mass. Likewise, the Universe was about 1 μs old when particle energies were comparable to the proton mass. Before this time, which was when quarks started binding into strongly-interacting particles, the history of the early Universe was dominated by elementary particles, such as quarks and leptons, and their fundamental interactions. Accelerators such as LEP have explored particle physics at energies up to about 100 GeV, corresponding to an age of about 10^{-10} s, and the physics suitable for describing the history of the Universe after this time is well understood. The LHC will explore particle energies up to about 1 TeV, corresponding to an age of about 10^{-12} s. Before this time, cosmology must rely on an extrapolation of the established laws of physics. This is, in particular, the case for inflation ⁷⁾, which is thought to have occurred when the Universe was younger than 10^{-30} s.

The abundance of supersymmetric dark matter would have been determined when the Universe was about 10^{-10} s, during a period for which LEP has measured, and the LHC will continue to measure, the microphysics. On the other hand, the abundance of superheavy cold dark matter would have been determined around the inflationary epoch, beyond the reach of accelerators.

However, inflationary cosmology is now being tested intensively by measurements of the cosmic microwave background (CMB) radiation ⁸⁾. These confirm that the overall energy density in the Universe today is very close to

the critical value $\Omega = 1$ predicted by inflation. The relative heights of the acoustic peaks imply that $\Omega_b \sim 0.05$, in agreement with Big-Bang nucleosynthesis ⁹⁾. The combination of CMB and large-scale structure data indicate that the total mass density $\Omega_m \simeq 0.3$. This is supported by data on high- z supernovae ¹⁰⁾, which also suggest that the present-day Universe is dominated by vacuum energy: $\Omega_\Lambda \simeq 0.65$ ¹¹⁾.

In later sections of these lectures, we discuss candidates for the missing dark matter: $\Omega_{CDM} \simeq \Omega_m - \Omega_b$. The explanation of Ω_Λ is a major challenge for a TOE, but one not discussed in these lectures.

1.2 Why Supersymmetry?

As already mentioned, in the Standard Model particle masses are believed to be due to a Higgs boson. The precision electroweak data from LEP and elsewhere are consistent with a relatively light Higgs boson weighing close to the present experimental limit of 114 GeV ¹²⁾. The next issue is why the W , Higgs and other particle masses are so much smaller than $m_P \equiv 1/\sqrt{G_N} \simeq 10^{19}$ GeV, the energy scale at which Einstein gravity would become as strong as the other particle forces. Alternatively, one may ask why is $G_F \sim 1/m_W^2 \gg G_N$? or why is the Coulomb potential in an atom so much larger than the Newton potential: $e^2 \gg G_N m_e m_A$?

It is not sufficient merely to set up this hierarchy and forget about the problem, because quantum effects make big corrections to electroweak boson masses:

$$\delta m_{H,W}^2 = \mathcal{O}\left(\frac{\alpha}{\pi}\right) \times \Lambda^2, \quad (2)$$

where Λ is some energy cut-off scale at which the Standard Model breaks down. If $\Lambda \sim m_{GUT}$ or m_P , the corrections (2) are many orders of magnitude larger than the physical values $m_{H,W} \sim 100$ GeV. Obtaining the physical value would require some apparently unnatural cancellation to many decimal places between the input classical values of the boson masses and the corrections (2). However, the latter may be made naturally small by postulating a supersymmetric theory ¹⁾, i.e., one with equal numbers of bosons and fermions with identical

couplings:

$$\delta m_{H,W}^2 = \mathcal{O}\left(\frac{\alpha}{\pi}\right) (m_B^2 - m_F^2). \quad (3)$$

The corrections (3) are comparable with the physical values if the supersymmetric partners have similar masses:

$$|m_B^2 - m_F^2| \lesssim 1 \text{TeV}^2. \quad (4)$$

There are other theoretical motivations for supersymmetry, but this is the only clear phenomenological indication that supersymmetry should appear at low energies.

Circumstantial support for this hypothesis is provided by the agreement between the values of the gauge couplings measured at low energies and the prediction of a GUT with low-energy supersymmetry ¹³⁾. Moreover, the lightest Higgs boson in the minimal supersymmetric extension of the Standard Model (MSSM) is predicted to weigh $\lesssim 130 \text{ GeV}$ ¹⁴⁾, in agreement with the indication from precision electroweak data that $m_H \lesssim 200 \text{ GeV}$ ¹⁵⁾.

In the minimal supersymmetric Standard Model (MSSM) ¹⁶⁾, one postulates a doubling of the familiar particle spectrum with spins differing by half a unit: $\gamma + \tilde{\gamma}, W + \tilde{W}, Z + \tilde{Z}, g + \tilde{g}, \ell + \tilde{\ell}, q + \tilde{q}$ and $H + \tilde{H}$ (moreover, supersymmetry requires two Higgs doublets H). Supersymmetry guarantees equal couplings for the spartner particles, but the sparticle masses are unknown apart from the expectation that they be $\lesssim 1 \text{ TeV}$. In this lecture, for simplicity we postulate universal sclar masses m_0 for the $\tilde{\ell}, \tilde{q}$ and H before renormalization, and likewise equal fermion masses for the different gauginos \tilde{g} etc.. The most important other parameters of this constrained MSSM (CMSSM) are $\tan \beta$, the ratio of Higgs vacuum expectation values, and μ , the mixing between the Higgs multiplets. The latter is fixed, up to a sign, by requiring electroweak symmetry breaking (EWSB).

In many supersymmetric models, the lightest supersymmetric particle (LSP) is expected to be stable, and hence present in the Universe today as a relic from the Big Bang ²⁾. The stability of the LSP follows from the conservation of a multiplicative quantum number called R parity, which takes the values $+1$ for all conventional particles and -1 for all supersymmetric particles. Its conservation is related to that of baryon number B , lepton number L and

spin S : $R = (-1)^{3B+L+2S}$. The conservation of R parity has three important consequences: (a) sparticles are always produced (or annihilate) in pairs, (b) heavier sparticles decay into lighter ones, and (c) the LSP is stable because it has no legal decay modes.

The LSP cannot have electromagnetic charge or strong interactions²⁾. If it did, it would dissipate energy during the formation of structures in the Universe, and condense along with ordinary matter, forming anomalous heavy isotopes. However, the upper limits on the abundances of such isotopes are many orders of magnitude below calculations of the relic LSP abundance. We infer that the LSP does not bind to ordinary nuclei, and hence can have no strong or electromagnetic interactions.

Candidates for the LSP include the sneutrinos of spin 0, the lightest neutralino χ – a combination of the $\tilde{\gamma}$, \tilde{H}^0 and \tilde{Z}^0 , with spin $\frac{1}{2}$, and the gravitino – the superpartner of the graviton, with spin $\frac{3}{2}$. A sneutrino LSP is excluded by a combination of Z^0 decay data from LEP and direct dark matter searches, whilst the gravitino is generally heavier than the lightest neutralino χ in models based on supergravity, as we assume here.

1.3 Constraints on Supersymmetry

There are important experimental constraints on supersymmetry from LEP¹⁷⁾, which has established a lower limit $m_{\chi^\pm} \gtrsim 103$ GeV on the mass of the lighter chargino (partner of the W^\pm and H^\pm), $m_{\tilde{e}} \gtrsim 99$ GeV and $m_H > 114.1$ GeV for the Standard Model Higgs boson. The latter also applies in the general MSSM if $\tan\beta \gtrsim 8$, and in the CMSSM for essentially all the $\tan\beta$ values of interest. This limit constrains indirectly the stop mass, which contributes a large loop correction to m_h :

$$\delta m_h^2 \sim \frac{m_t^4}{m_W^2} \ln \left(\frac{m_{\tilde{t}}^2}{m_t^2} \right). \quad (5)$$

In the CMSSM, this may be translated into a lower limit on $m_{1/2}$, as illustrated in Fig. 1. There are also direct limits on squark and gluino masses from LEP and the Tevatron that are not shown there.

Another important constraint is provided by the fact that the rare B

meson decay $b \rightarrow s\gamma$ has been observed at a level close to its Standard Model value¹⁸⁾. This constraint is particularly important for $\mu < 0$ and for large $\tan\beta$, as seen in Fig. 1. Some recent interest has been sparked by the recent indication that the anomalous magnetic moment of the muon, $g_\mu - 2$ ¹⁹⁾, may differ from the Standard Model prediction. However, the interpretation of the data is not yet clear, and we treat this here as an optional constraint.

The final constraint shown in Fig. 1 is that due to the supersymmetric relic density¹⁷⁾. Generically, the mass density $\rho_\chi \equiv m_\chi n_\chi$, where the number density n_χ is inversely proportional to the annihilation cross section: $n_\chi \propto 1/\sigma_{ann}(\chi\chi \rightarrow \dots)$. Since a typical annihilation cross section $\propto 1/m_\chi^2$, the relic density typically increases with the mass, and $\Omega_\chi h^2 < \Omega_{CDM} \lesssim 0.3$ only for $m_\chi \lesssim 1$ TeV. However, there are filaments of parameter space extending to larger m_χ . For example, when the next-to-lightest supersymmetric particle (NLSP) \tilde{X} is close in mass to the LSP, coannihilation processes $\chi\tilde{X} \rightarrow \dots$ may also reduce the relic density²⁰⁾, and hence increase the upper limit on m_χ . This effect is visible at large $m_{1/2}$ in Fig. 1, where $\chi\tilde{\tau}$ coannihilation is important. Sometimes also a direct-channel resonance may enhance the annihilation rate along some ‘funnel’ in parameter space²¹⁾, as also seen in Fig. 1b at large m_0 ¹⁷⁾.

A set of benchmark supersymmetric models has been proposed²²⁾, taking into account the constraints from LEP, $b \rightarrow s\gamma$, the cosmological relic density and (optionally) $g_\mu - 2$. As seen in Fig. 2a, the points were chosen to illustrate the range of supersymmetric possibilities, rather than to sample the parameter space ‘fairly’ in a statistical sense. As also seen in Fig. 2a, five of the proposed points are in the ‘bulk’ of the cosmologically-preferred region at small $m_{1/2}$ and m_0 , four are spread along the coannihilation ‘tails’ towards large $m_{1/2}$, two are in the ‘focus-point’ region at large m_0 , and two are along rapid-annihilation ‘funnels’ at large $m_{1/2}$ and m_0 . As seen in Fig. 2b, the various points proposed have $\tan\beta = 5, 10, 20, 35$ and 50 . Most have $\mu > 0$, as favoured by $g_\mu - 2$, but two have $\mu < 0$. About a half of the points yield values of $g_\mu - 2$ within two standard deviations of the present central experimental value.

The benchmark points were first used to discuss the supersymmetric physics reaches of different particle accelerators²²⁾. Here they are used to an-

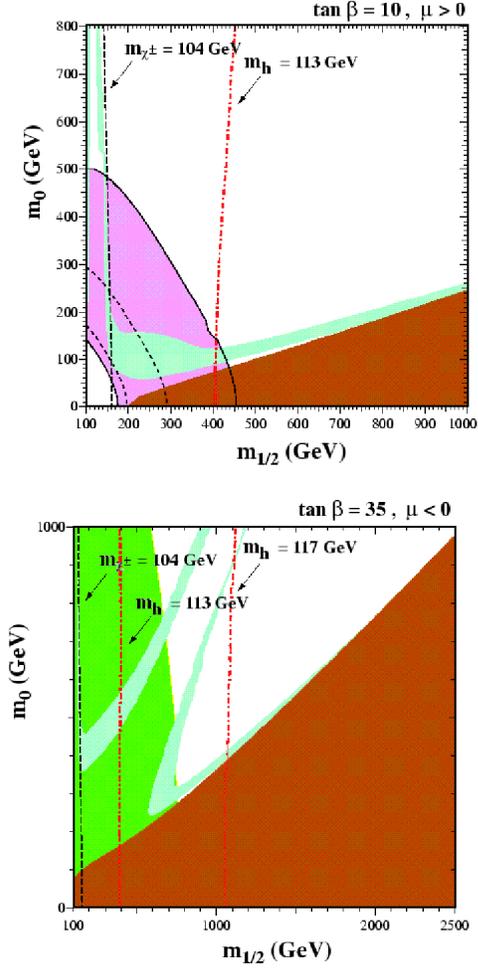


Figure 1: The $m_{1/2}, m_0$ plane for the CMSSM with (a) $\tan\beta = 10$, $A = 0$ and $\mu > 0$, and (b) $\tan\beta = 35$, $A = 0$ and $\mu < 0$, showing the region preferred by the cosmological relic density constraint $0.1 \leq \Omega_\chi h^2 \leq 0.3$ (light, green shading), the excluded region where $m_{\tilde{\tau}} < m_\chi$ (dark, brown shading), and the region disallowed by $b \rightarrow s\gamma$ (dark, green shading) ¹⁷. The region preferred by $g_\mu - 2$ in (a) is shown as a (medium, pink) shaded diagonal band. Also shown as a near-vertical line is the contour $m_h = 113$ GeV for $m_t = 175$ GeV, which provides a lower limit on $m_{1/2}$. For comparison, we also exhibit the reaches of LEP 2 searches for charginos χ^\pm .

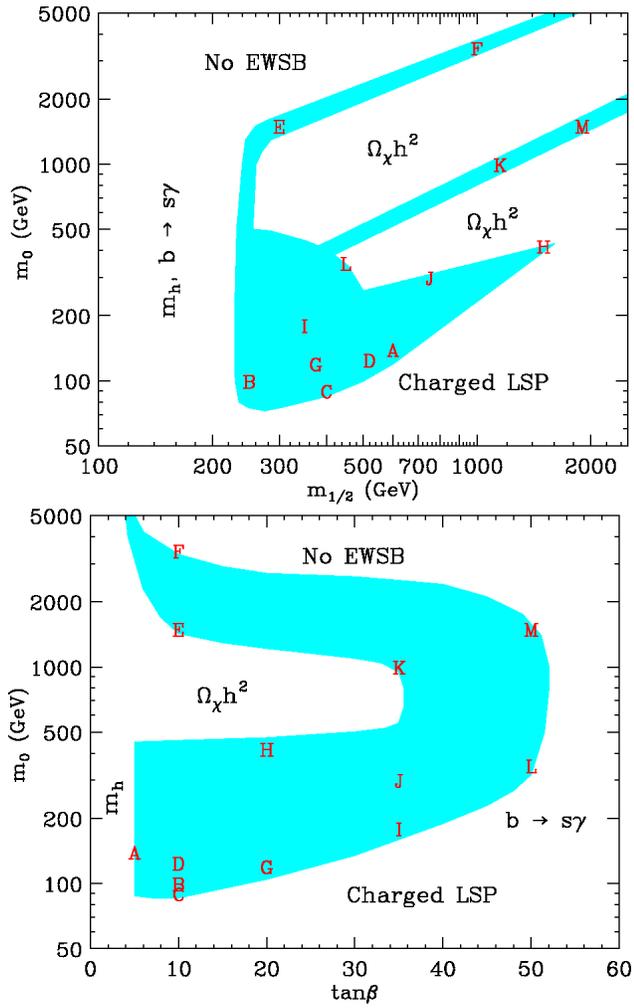


Figure 2: Locations of the proposed CMSSM benchmark points ²²⁾ in (a) the $(m_{1/2}, m_0)$ plane, and (b) the $(\tan\beta, m_0)$ plane. The shaded areas roughly indicate the various cosmologically preferred regions discussed in the text.

alyze the sensitivities of different non-accelerator searches for supersymmetric dark matter ²³⁾.

1.4 Searches for Supersymmetric Dark Matter

One strategy is to look for the products of the annihilations of supersymmetric relic particles *in the galactic halo*. Searches for cosmic-ray antiprotons ²⁴⁾, positrons and photons have been proposed. Measurements of the cosmic-ray antiproton flux agree with calculations of secondary production by primary matter cosmic rays, and exhibit the expected modulation during the solar cycle ²⁵⁾. Moreover, modern predictions of the flux due to relic annihilations are not very optimistic, so we do not pursue further this possibility. Observations of cosmic-ray positrons exhibit a possible excess at energies around 10 GeV, but the benchmark scenarios predict much smaller fluxes ²³⁾.

The flux of gamma rays due to relic annihilations may be enhanced in the direction of *the galactic centre* by a factor $J \lesssim 10^5$. We show in Fig. 3 predictions for the benchmark scenarios assuming $J = 500$, compared with the sensitivities of terrestrial and space experiments ²³⁾. We see that models *I* and *L* may offer the best prospects, particularly for a low-threshold detector such as GLAST. However, within the astrophysical uncertainties, several other models offer prospects, including models *K, E, B, G, F, M* and *J* in particular.

Another possibility is to look for energetic muons produced by energetic neutrinos from relic annihilations in *the core of the Sun or Earth* ²⁶⁾. According to our calculations ²³⁾, the Sun offers better prospects in our benchmark scenarios, particularly in models *E, B* and *F*, as seen in Fig. 4.

Finally, we consider the prospects for *direct detection* of supersymmetric dark matter via its elastic scattering on nuclei in an underground laboratory ²⁷⁾. This scattering has both a spin-dependent component, which is related to the different quark contributions to the nucleon spin, and a spin-independent contribution, related to the different quark contributions to the nucleon mass ²⁸⁾. Within our benchmark scenarios, the latter offer better prospects to planned detectors ²³⁾. As seen in Fig. 5, the proposed GENIUS experiment has good prospects in models *B, E, I, G, L* and *F*.

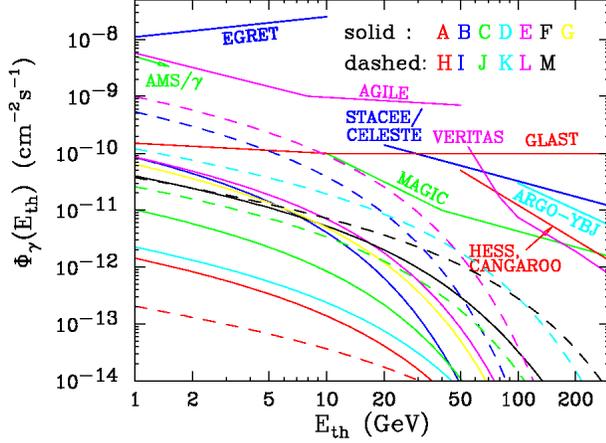


Figure 3: The integrated photon flux $\Phi_\gamma(E_{thr})$ as a function of photon energy threshold E_{thr} for photons produced by relic annihilations in the galactic center ²³). A moderate halo enhancement parameter $J = 500$ is assumed. Point source flux sensitivities for various gamma ray detectors are also shown.

Combining all these analyses, we see that non-accelerator experiments have good prospects for detecting quite a large proportion of the proposed benchmark models.

2 Superheavy Relic Particles

2.1 Ultrahigh-Energy Cosmic Rays

The flux of cosmic rays falls approximately as E^{-3} from $E \sim 1$ GeV, through $E \sim 10^6$ GeV where there is a small change in slope called the ‘knee’, continuing to about 10^{10} GeV, the ‘ankle’. Beyond about 5×10^{10} GeV, as seen in Fig. 6, one expects a cutoff due to the photopion reaction $p + \gamma_{CMB} \rightarrow \Delta^+ \rightarrow p + \pi^0, n + \pi^+$, for all primary cosmic rays that originate from more than about 50 Mpc away ²⁹). However, several experiments see cosmic-ray events with higher energies of 10^{11} GeV or more ³⁰). If this excess flux beyond the GZK cutoff is confirmed, conventional physics would require it to originate from

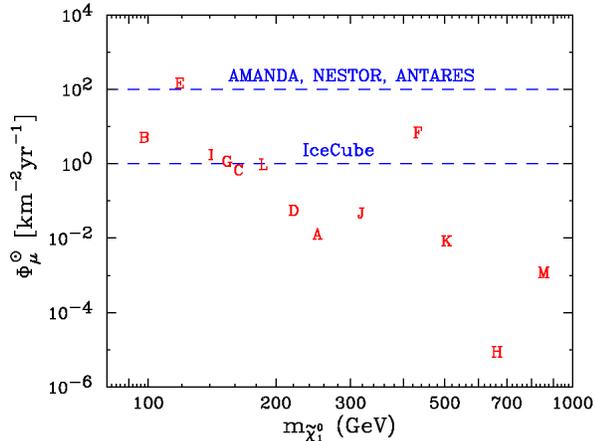


Figure 4: *Muon fluxes from neutrinos originating from relic annihilations inside the Sun* ²³). Approximate sensitivities of near future neutrino telescopes ($\Phi_\mu = 10^2/\text{km}^2/\text{y}$ for AMANDA II, NESTOR, and ANTARES, and $\Phi_\mu = 1/\text{km}^2/\text{y}$ for IceCube) are also indicated.

distances $\lesssim 100$ Mpc, in which case one would expect to see some discrete sources. Analogous cutoffs are expected for primary cosmic-ray photons or nuclei, as also seen in Fig. 6.

There are two general categories of sources considered for ultrahigh-energy cosmic rays (UHECRs): *bottom-up* and *top-down* scenarios.

Astrophysical sources capable of accelerating high-energy cosmic rays must be larger than the gyromagnetic radius R corresponding to their internal magnetic field B :

$$R \sim \left(\frac{100}{Z}\right) \left(\frac{E}{10^{11}\text{GeV}}\right) \left(\frac{\mu G}{B}\right) \text{ kpc}, \quad (6)$$

where Z is the atomic number of the cosmic ray particle. Candidate astrophysical sources include gamma-ray busters (GRBs) and active galactic nuclei (AGNs).

If UHECRs are produced by such localized sources, one would expect to see a clustering in their arrival directions. Such clustering has been claimed

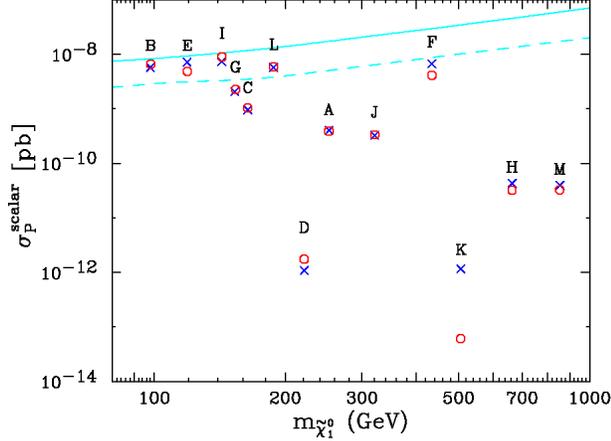


Figure 5: *Elastic cross sections for spin-independent neutralino-proton scattering, using two different codes (blue crosses and red circles) are compared. Projected sensitivities for CDMS II and CRESST (solid) and GENIUS (dashed) are also shown ²³⁾.*

in both the AGASA and Yakutsk data ³¹⁾, but I personally do not find the evidence overwhelming. A correlation has also been claimed with BL Lac objects ³²⁾, which are AGNs emitting relativistic jets pointing towards us, but this is also a claim that I should like to see confirmed by more data, as will be provided soon by the HiRes and Auger ³³⁾ experiments.

Favoured top-down scenarios involve physics at the GUT scale $\gtrsim 10^{15}$ GeV that produces UHECRs with energies $\sim 10^{12}$ GeV via some ‘trickle-down’ decay mechanism. Suggestions have included topological defects, such as cosmic strings that radiate energetic particles, and the decays of metastable superheavy relic particles ⁴⁾.

In the latter case, one would expect most of the observed UHECRs to come from the decays of relics in our own galactic halo. In this case, one would expect the UHECRs to exhibit an anisotropy correlated with the orientation of the galaxy. The present data are insufficient to confirm or exclude an isotropy of the magnitude predicted in different halo models, but the Auger experi-

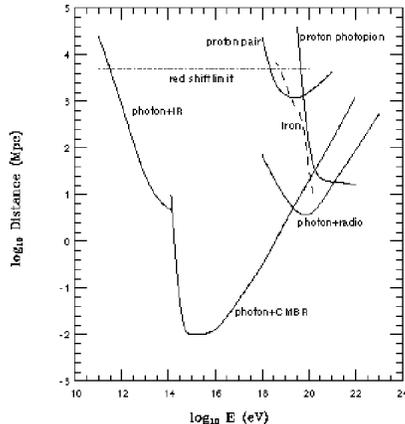


Figure 6: *The cut-offs expected for high-energy protons ²⁹⁾, Iron nuclei and photons, due to photo-absorption processes and e^+e^- pair production, respectively ³⁰⁾.*

ment ³³⁾ should be able to decide the issue ³⁴⁾. One might naively expect that superheavy relic particles would be spread smoothly throughout the halo, and hence that they would not cause clustering in the UHECRs. However, this is not necessarily the case, as many cold dark matter models predict clumps within the halo, which could contribute a clustered component on top of an apparently smooth background ³⁵⁾.

2.2 Metastable Superheavy Relic Particles

The proton is a prototype for a metastable massive particle. We know that its lifetime must exceed about 10^{33} y, much longer than if it decayed via conventional weak interactions, but there is no known exact symmetry principle capable of preventing the proton from decaying. We believe that it is only metastable, decaying very slowly via some higher-dimensional non-renormalizable interaction that violates baryon number. For example, in many GUT models there is a dimension-6 $qqq\ell$ interaction with a coefficient $\propto 1/M^2$, where M is some superheavy mass scale. This would yield a decay amplitude

$A \sim 1/M^2$, and hence a lifetime

$$\tau = \frac{1}{\Gamma} \sim \frac{|A|^{-2}}{m_p^5} \sim \frac{M^4}{m_p^5}. \quad (7)$$

This estimate yields a lifetime $\gtrsim 10^{33}$ y if $M \gtrsim 10^{15}$ GeV.

We must work harder in the case of a superheavy relic weighing $\gtrsim 10^{12}$ GeV, but the principle is the same. For an interaction of dimension $4 + n$, we expect

$$\tau \sim \frac{M^{2n}}{m_{\text{relic}}^{2n+1}} \quad (8)$$

This could yield a lifetime greater than the age of the Universe, even for $m_{\text{relic}} \sim 10^{12}$ GeV, if M and/or n are large enough, for example if $M \sim 10^{17}$ GeV and $n \geq 9$ 36).

Phenomenological constraints on such metastable relic particles were considered some time ago for reasons other than explaining UHECRs 37). Constraints from the abundances of light elements, from the CMB and from the high-energy ν flux have been considered. They provide no obstacle to postulating a superheavy relic particle with $\Omega h^2 \sim 0.3$ if $\tau \gtrsim 10^{15}$ y. Hence, metastable superheavy relic particles could in principle constitute most of the cold dark matter.

Possible theoretical candidates within a general framework of string and/or M theory have been considered 36). There are several possible origins for massive states in such a theory. For example, there are Kaluza-Klein states - ‘hexons’ - that appear when one compactifies six surplus dimensions, reducing the number of string theory dimensions from 10 to 4, or M theory from 11 dimensions to 5. However, in most models these hexons are highly unstable. Then, in M theory there are massive states - ‘pentons’ - that appear upon compactification from 5 to 4 dimensions, but these too are expected to be unstable. The most plausible candidates appear to be ‘cryptons’ 4), massive states from the so-called ‘hidden sector’ of string theory or M theory, which do not share gauge interactions with conventional matter.

String models have the generic feature that, in addition to the gauge interactions of the Standard Model, there are others that act on a different set

of such ‘hidden’ matter particles. Just as our gauge interactions bind quarks to form metastable massive particles, the protons, so the ‘hidden-sector’ gauge interactions might become strong at some very high energy scale, and form analogous, but supermassive, metastable particles. Indeed, such bound states avoid the appearance of fractionally-charged particles in generic string models 4).

Just like the proton, these massive cryptons generally decay through high-dimension interactions into multiple quarks and leptons. The energetic quarks hadronize via QCD in a way that can be modelled using information from Z^0 decays at LEP. Several simulations have shown that the resulting spectrum of UHECRs is compatible with the available data, whether supersymmetry is included in the jet fragmentation process, or not: one example is shown in Fig. 7 38).

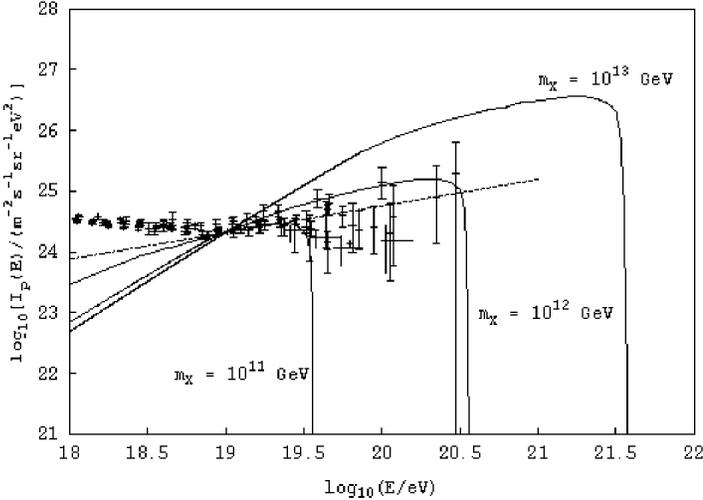


Figure 7: *The observed spectrum of ultra-high-energy cosmic rays, which is compatible with a calculation of crypton decays, for various different choices of the crypton mass 38).*

2.3 Abundance of Superheavy Relic Particles

A crucial issue is whether there is a mechanism that might produce a relic density of superheavy particles that is large enough to be of interest for cosmology, without being excessive. As was discussed in the first lecture, the plausible upper limit on the mass of a relic particle that was initially in thermal equilibrium is of the order of a TeV. However, equilibrium might have been violated in the early Universe, around the epoch of inflation, and various non-thermal production mechanisms have been proposed ³⁹⁾. These include out-of-equilibrium processes at the end of the inflationary epoch, such as parametric resonance effects, and gravitational production as the scale factor of the Universe changes rapidly. Calculations with some particular examples of inflationary potentials indicate that superheavy relic particles might be produced with a significant fraction of the critical density.

2.4 Prospects

We have seen that UHECRs could perhaps be due to the decays of metastable superheavy relic particles. They might have the appropriate abundance, their lifetimes might be long on a cosmological time-scale, and the decay spectrum might be compatible with the events seen. Pressure points on this interpretation of UHECRs include the composition of the UHECRs – there should be photons and possibly neutrinos, as well as protons, and no heavier nuclei; their isotropy – UHECRs from relic decays would exhibit a detectable galactic anisotropy; and clustering – this would certainly be expected in astrophysical source models, but is not excluded in the superheavy relic interpretation.

The Auger project ³³⁾ currently under construction in Argentina should provide much greater statistics on UHECRs and be able to address many of these issues. In the longer term, the EUSO project ⁴⁰⁾ now being considered by ESA for installation on the International Space Station would provide even greater sensitivity to UHECRs. Thus an experimental programme exists in outline that is capable of clarifying their nature and origin, telling us whether they are indeed due to new fundamental physics.

3 Gamma-Ray Busters and Quantum Gravity

3.1 Introduction

Later at this school, there are several lectures on gamma-ray busters (GRBs) and their associated astrophysics. In this lecture, I briefly preview a few aspects that are relevant to our suggestion that they may be used to test fundamental physics, specifically some speculations about quantum gravity.

The discovery of GRBs by the Vela satellites – launched to search for nuclear tests in space – was announced in 1973. The study of GRBs was revolutionized by the launch in 1991 of the Compton Gamma-Ray Observatory (CGRO), which observed about 3000 GRBs, at a rate of about one per day, until it was de-orbited. GRBs were observed to have an essentially isotropic angular distribution, suggesting that they had a cosmological origin – one of the key ingredients in their suitability for testing quantum gravity. The γ rays were observed to have non-thermal spectra of the generic form $N(E) \sim E^{-\alpha}$: $\alpha \sim 1$ to 3. Thus there are more high-energy photons than would be expected from a purely thermal spectrum. Subsequently, there have been reported detections of GRBs in the GeV range, e.g., by EGRET ⁴¹⁾, and even of TeV γ rays, by MILAGRITO ⁴²⁾. Such high-energy emissions are of particular interest for testing quantum gravity, as we see later. Also important is the fact that GRB emissions exhibit time structures as short as 10^{-3} s. This suggests that the core of the progenitor might be as small as about 10 km, which is important for astrophysical models. This time structure also provides the lever for probing quantum gravity, as also discussed later ⁴³⁾.

The second revolution in GRB observations was triggered by the Beppo-SAX satellite, which was able to observe X-ray afterglows from GRBs. These enabled accurate directions to be established, which in turn permitted optical and radio detections, and the identification of host galaxies. These observations in turn permitted the redshifts of the GRBs to be measured for the first time. Those measured have $z = \mathcal{O}(1)$, confirming the earlier expectation that they must be at cosmological distances. However, the GRBs with measured redshifts had bursts of relatively long duration. It is thought that there might be a separate population of short-duration GRBs, whose progenitors might be

systematically different, perhaps with a distinct redshift distribution.

3.2 Models of GRBs

As already mentioned, the time structures observed in GRBs imply that the cores of their progenitors should have a size ~ 10 km⁴⁴). To be seen at redshifts $z \sim 1$ with the observed strengths, they must emit a total energy $\sim 10^{53}$ ergs, comparable to the binding energy of a neutron star, *if the burst is isotropic*. If not, and there is increasing evidence that (at least some) GRBs emit beams of radiation, then the energy requirement is reduced by a factor $\Delta\Omega/4\pi$, but is nevertheless substantial. The non-thermal spectrum can be explained if the GRB progenitor produces a relativistic fireball with $\gamma \sim 10^2$. The γ emission is generally thought to originate from internal shocks due to irregularities in the velocity of the relativistic flow. These convert its kinetic energy into radiation energy, whose escape requires the medium to be rather transparent. Eventually, the relativistic flow slows as it ploughs into the interstellar medium, producing an external shock, and reverse shocks are also expected. In this way, the original γ radiation is succeeded by X rays, ultraviolet, optical, infrared and radio emission. Observations of radio scintillation confirm that the size of the source is as expected from an initially relativistic expansion.

It has been suggested that GRBs might be important sources of UHECRs, since they seem to meet the requirements for magnetic fields strong enough to confine protons while they are accelerated to $\sim 10^{11}$ GeV, and they are believed to produce relativistic shocks. Equipartition of energy between electrons and protons would yield a cosmic-ray energy density of $\sim 10^{44}$ ergs/Mpc³/y, which is comparable to the observed energy density of cosmic rays. Simulations assuming a homogeneous cosmological distribution of GRB sources, each with an energy spectrum

$$\frac{du}{dE_p} \sim E_p^{-2}, \quad (9)$$

with the total energy equal to the electron energy in the GRB fireballs show that a GRB origin for UHECRs cannot apparently be excluded on the basis of energetics.

It is interesting to note that a mechanism of this type would also yield a

spectrum of UHE neutrinos, from reactions such as $p + \gamma \rightarrow n + \pi^\pm, \mu^\pm \rightarrow \nu$. If these also extend up to 10^{10} GeV, as suggested by some authors ⁴⁵⁾, they would provide an ideal way to test the violation of Lorentz invariance suggested in some speculations on space-time foam ⁴⁷⁾, as we now discuss.

3.3 Violation of Lorentz Invariance?

Space-time is well known to be approximately flat at large distance scales. However, one expects large energy and topology fluctuations $\Delta E, \Delta\chi$ at short distance scales Δx , over short time periods Δt :

$$\Delta E \sim 10^{19} \text{ GeV}, \quad \Delta\chi \sim 1 \quad \text{in} \quad \Delta x \sim \ell_P \sim 10^{-33} \text{ cm}, \quad \Delta t \sim t_P \sim 10^{-43} \text{ s}. \quad (10)$$

Are there any observable consequences of such quantum-gravitational fluctuations – ‘space-time foam’? One possibility might be microscopic loss of quantum coherence, as information leaks away through microscopic event horizons ⁴⁸⁾. Another suggested possibility is that special relativity might be modified, with the velocity of light becoming energy-dependent:

$$C(E) \simeq c \left(1 - \frac{E}{M_{QG}} + \dots \right) \quad (11)$$

where M_{QG} might be $\mathcal{O}(m_P) \simeq 10^{19}$ GeV ⁴³⁾.

One way to motivate this latter suggestion is to imagine balls of different sizes (corresponding to photons of different wavelengths, i.e., frequencies) rolling across a rough surface. Balls that are much larger than the corrugations in the surface will ignore them, whereas small balls (high-frequency photons) will roll up and down the corrugations, taking a longer time to propagate a long (astrophysical) distance. The vacuum would appear to have a non-trivial refractive index.

This intuition has been modelled, in particular, in a D -brane model of space-time foam ⁵⁾, where the interactions of particles with solitonic ‘lumps’ result in a reduction in velocity by a factor $(1 - E/M)$, where M is a characteristic mass-scale for the ‘lumps’. Analogous effects on photon propagation have been found in other models of space-time foam. For example, one possibility suggested by the loop-gravity approach is birefringence, i.e., a difference

between the velocities of high-energy photons of different helicities ⁴⁹⁾. It has also been pointed out that light-cone fluctuations could modify the propagation of photons and other energetic particles ⁵⁰⁾.

In all these approaches, the effect on particle propagation *increases* with its energy, whereas conventional in-medium effects tend to be suppressed at high energies. This is the case, for example, for photon through a finite-density or -temperature plasma.

3.4 Astrophysical Probes of Lorentz Violation

Since the propagation time for a photon coming from a source at distance D is $t = D/v$, where the velocity $v \neq c$ in general, the time delay due to a reduction in velocity Δv is $\Delta t \simeq (D/v^2)\Delta v$. Assuming $v \simeq c(1 - E/M_{QG})$, we find

$$\Delta t \simeq \frac{D}{c^2} \frac{E}{M_{QG}} \quad (12)$$

The figure of merit for astrophysical sensitivity is therefore

$$M_{QG} \simeq \frac{E \cdot D}{\Delta t}, \quad (13)$$

where Δt is the available time resolution. Clearly there is a premium on high-energy sources at large distances D , with short time structures Δt in their emissions.

Astrophysical sources that have been considered include GRBs, AGNs and pulsars: representative estimates of their sensitivities (13) are shown in Table 1 ⁶⁾. We see that they approach within just a few orders of magnitude of m_P , the ball-park scale where one expect to find m_{QG} , if this type of Lorentz violation appears at all.

We made a systematic analysis of all the GRBs with cosmological redshifts that had been measured some time ago ⁶⁾. We analyzed the time structures in BATSE and OSSE data, comparing the arrival time series of photons in the lowest and highest energy channels. We made fits using several different functional forms for the peaks, as seen in Fig. 8, so as to control one possible

Source	Distance	E	Δt	Sensitivity to M
GRB 920229	3000 Mpc (?)	200 keV	10^{-2} s	0.6×10^{16} GeV (?)
GRB 980425 ^a	40 Mpc	1.8 MeV	10^{-3} s (?)	0.7×10^{16} GeV (?)
GRB 920925c ^a	40 Mpc (?)	200 TeV (?)	200 s	0.4×10^{19} GeV (?)
Mrk 421	100 Mpc	2 TeV	280 s	$> 7 \times 10^{16}$ GeV
Crab pulsar	2.2 kpc	2 GeV	0.35 ms	$> 1.3 \times 10^{16}$ GeV
GRB 990123	5000 Mpc	4 MeV	1 s (?)	2×10^{16} GeV (?)

Table 1: *The mass-scale parameter M_{QG} is defined by $\delta v/c = E/M_{QG}$. The question marks in the Table indicate uncertain observational inputs. Hard limits are indicated by inequality signs. This Table is adapted from ⁶⁾, where original references may be found.*

source of systematic error. We then compared the arrival times and widths of the peaks in different energy bins. Since differences in these quantities could in principle be generated at the source, we looked for a possible correlation with the propagation distance, as determined by the measured redshift, as seen in Fig. 9. The data were completely consistent with the absence of any such correlation, and we concluded that

$$M_{QG} > 10^{15} \text{ GeV} \quad (14)$$

We regard this as a statistically and systematically secure conclusion, that does not rely on the interpretation of any individual GRB event.

We expect the redshifts of many more GRBs to be measured with the advent of early-warning satellites such as HETE-II. We also expect increased sensitivity to higher-energy photons, with the way being led by the MILAGRITO report of a possible TeV photon signal from GRB 970417a ⁴²⁾. Space experiments such as AGILE, AMS and GLAST may play important rôles, and there have been some preliminary analyses of their prospective sensitivities that seem quite encouraging ⁴⁶⁾.

Another exciting experimental possibility is to look for high-energy ν pulses from GRBs. In some models, neutrinos with energies up to 10^{11} GeV may be emitted. If a pulse were seen in coincidence with a GRB signal, it would provide sensitivity to $M_{QG} \gtrsim 10^{26}$ GeV! Conversely, if $M_{QG} \sim 10^{19}$ GeV, any such neutrino pulse would be spread over a period of years, becoming unobservable ⁴⁷⁾!

GRB 990123: BATSE data Ch. 1 and Ch. 3

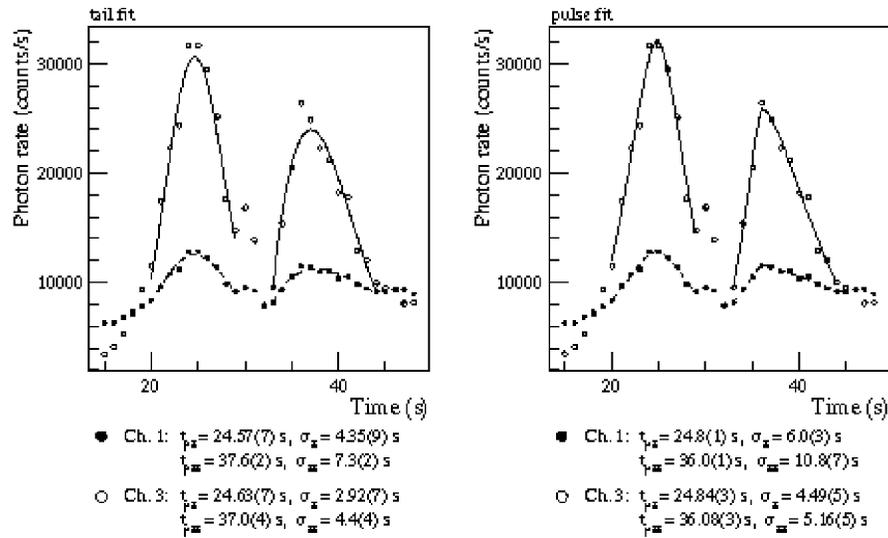


Figure 8: Fits to the pulses of GRB 990123 in different BATSE energy channels, using two different fitting functions ⁶).

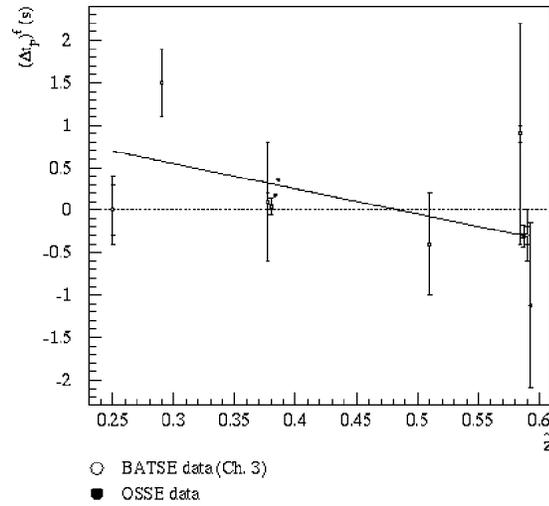


Figure 9: A fit to the time-lags extracted from the pulses of GRBs with measured redshifts. No significant dependence on the redshift was found ⁶⁾.

3.5 Erasure of the GZK Cutoff?

We have already discussed in connection with UHECRs the fact that the Universe should be opaque to UHE protons over distances $\gtrsim 100$ Mpc, at energies $\gtrsim 10^{11}$ GeV, due to the reaction $p + \gamma_{CMB} \rightarrow N + \pi$. It has been pointed ⁵¹⁾ out that this process is forbidden at high energies if conventional relativistic kinematics is modified by a dispersion relation of the form

$$C^2 \cdot p^2 = E^2 \left(1 + \frac{E}{M} + \dots \right) \quad (15)$$

as discussed above. In the absence of particle production, there would be no GZK cutoff.

According to conventional physics, there should be an analogous cutoff for photons, as seen in Fig. 6, due to pair production in the infrared background: $\gamma + \gamma_{IR} \rightarrow e^+ e^-$. In the particular case of the AGN Mkn 501, the photon energy spectrum should be cut off above 1 TeV ⁵²⁾. However, higher-energy photons have been reported by the HEGRA collaboration ⁵³⁾. If one assumes that they are the attenuated residue of a real flux from Mkn 501, the inferred produced flux would have been extremely high at energies above 1 TeV. Many production mechanisms would predict similar fluxes of high-energy photons and neutrinos. The AMANDA collaboration has searched for TeV neutrinos from Mkn 501, and not seen a signal, excluding such a high neutrino flux ⁵⁴⁾. The interpretation is ambiguous: could HEGRA have calibrated incorrectly their observed photon energies? Could one produce high-energy photons without corresponding neutrinos? Might the IR background be smaller than thought? All these – and other – conservative hypotheses should be investigated and excluded before embracing the radical hypothesis of a violation of Lorentz invariance.

In connection with this last possibility, a second avenue for the modification of Lorentz kinematics should be kept in mind, as well as the modified dispersion relation. It has been argued that energy is conserved only statistically, and the possibility of energy non-conservation during particle interactions should be taken into account. The corresponding modification of the GZK cutoff could be comparable to that induced by the modified dispersion relation ⁵⁵⁾.

4 Conclusions

In these lectures, a very incomplete set of topics at the boundaries between astrophysics, cosmology, cosmic-ray physics and particle physics have been discussed. Nevertheless, they serve to illustrate the breadth and interest of this burgeoning field. It is very difficult to predict what will be the next great discovery in astroparticle physics, and it may well have no relation to the topics selected here. However, I am sure that astroparticle physics will continue to excite us for the foreseeable future.

References

1. L. Maiani, *Proceedings of the 1979 Gif-sur-Yvette Summer School On Particle Physics*, 1; G. 't Hooft, in *Recent Developments in Gauge Theories, Proceedings of the Nato Advanced Study Institute, Cargese, 1979*, eds. G. 't Hooft *et al.*, (Plenum Press, NY, 1980); E. Witten, *Phys. Lett. B* **105** (1981) 267.
2. J. Ellis, J.S. Hagelin, D.V. Nanopoulos, K.A. Olive and M. Srednicki, *Nucl. Phys. B* **238** (1984) 453; see also H. Goldberg, *Phys. Rev. Lett.* **50** (1983) 1419.
3. N. Hayashida *et al.*, *Astrophys. J.* **522** (1999) 225 and references therein.
4. J. Ellis, J. L. Lopez and D. V. Nanopoulos, *Phys. Lett. B* **247** (1990) 257.
5. J. R. Ellis, N. E. Mavromatos and D. V. Nanopoulos, *Int. J. Mod. Phys. A* **13** (1998) 1059; *Gen. Rel. Grav.* **31** (2000) 1257; *Phys. Rev. D* **61** (2000) 027503; arXiv:gr-qc/9909085.
6. J. Ellis, K. Farakos, N. E. Mavromatos, V. A. Mitsou and D. V. Nanopoulos, *Astrophys. J.* **535** (2000) 139. This paper contains a statistical analysis and references to previous indicative limits.
7. K. A. Olive, *Phys. Rept.* **190** (1990) 307.
8. For a recent review, see: S. Church, A. Jaffe and L. Knox, arXiv:astro-ph/0111203.

9. For a recent review, see: K. A. Olive, G. Steigman and T. P. Walker, Phys. Rept. **333** (2000) 389 [arXiv:astro-ph/9905320].
10. A. G. Riess *et al.* [Supernova Search Team Collaboration], Astron. J. **116** (1998) 1009 [arXiv:astro-ph/9805201]; S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], Astrophys. J. **517** (1999) 565 [arXiv:astro-ph/9812133].
11. See, for example, N. Bahcall, J. P. Ostriker, S. Perlmutter and P. J. Steinhardt, Science **284** (1999) 1481.
12. LEP Higgs Working Group for Higgs boson searches, arXiv:hep-ex/0107029 and arXiv:hep-ex/0107030.
13. J. Ellis, S. Kelley and D. V. Nanopoulos, Phys. Lett. B **260** (1991) 131; U. Amaldi, W. de Boer and H. Furstenau, Phys. Lett. B **260** (1991) 447; C. Giunti, C. W. Kim and U. W. Lee, Mod. Phys. Lett. A **6** (1991) 1745.
14. J. Ellis, G. Ridolfi and F. Zwirner, Phys. Lett. B **257** (1991) 83; M.S. Berger, Phys. Rev. D **41** (1990) 225; Y. Okada, M. Yamaguchi and T. Yanagida, Prog. Theor. Phys. **85** (1991) 1; Phys. Lett. B **262** (1991) 54; H.E. Haber and R. Hempfling, Phys. Rev. Lett. **66** (1991) 1815; R. Barbieri, M. Frigeni and F. Caravaglios, Phys. Lett. B **258** (1991) 167; P.H. Chankowski, S. Pokorski and J. Rosiek, Phys. Lett. B **274** (1992) 191; J.R. Espinosa and M. Quirós, Phys. Lett. B **266** (1991) 389; J.L. Lopez and D.V. Nanopoulos, Phys. Lett. B **266** (1991) 397; M. Carena, K. Sasaki and C.E.M. Wagner, Nucl. Phys. B **381** (1992) 66; P.H. Chankowski, S. Pokorski and J. Rosiek, Phys. Lett. B **281** (1992) 100; Nucl. Phys. B **423** (1994) 437; D.M. Pierce, A. Papadopoulos and S.B. Johnson, Phys. Rev. Lett. **68** (1992) 3678; A. Brignole, Phys. Lett. B **281** (1992) 284; M. Drees and M.M. Nojiri, Phys. Rev. D **45** (1992) 2482; V. Barger, M.S. Berger and P. Ohmann, Phys. Rev. D **49** (1994) 4908; G.L. Kane, C. Kolda, L. Roszkowski and J.D. Wells, Phys. Rev. D **49** (1994) 6173; R. Hempfling and A.H. Hoang, Phys. Lett. B **331** (1994) 99; P. Langacker and N. Polonsky, Phys. Rev. D **50** (1994) 2199; H.E. Haber, R. Hempfling and A.H. Hoang, Zeit. für Phys. C **75** (1997) 539; S. Heinemeyer, W. Hollik and G. Weiglein, Comput. Phys. Commun. **124**, 76 (2000) [hep-ph/9812320].

15. LEP Collaborations, ALEPH, DELPHI, L3, OPAL, the LEP Electroweak Working Group and the SLD Heavy Flavour and Electroweak Groups, LEPEWWG/2001-01, available from <http://lepewwg.web.cern.ch/LEPEWWG/welcome.html>.
16. H. P. Nilles, Phys. Rept. **110** (1984) 1; H. E. Haber and G. L. Kane, Phys. Rept. **117** (1985) 75.
17. J. R. Ellis, G. Gani, D. V. Nanopoulos and K. A. Olive, Phys. Lett. B **502** (2001) 171 [arXiv:hep-ph/0009355]; J. Ellis, D. V. Nanopoulos and K. A. Olive, Phys. Lett. B **508** (2001) 65 [arXiv:hep-ph/0102331]; J. R. Ellis, T. Falk, G. Gani, K. A. Olive and M. Srednicki, Phys. Lett. B **510** (2001) 236 [arXiv:hep-ph/0102098].
18. M.S. Alam et al., [CLEO Collaboration], Phys. Rev. Lett. **74** (1995) 2885 as updated in S. Ahmed et al., CLEO CONF 99-10; K. Abe *et al.*, [Belle Collaboration], [arXiv:hep-ex/0107065]; L. Lista [BaBar Collaboration], [arXiv:hep-ex/0110010]; C. Degrossi, P. Gambino and G. F. Giudice, JHEP **0012** (2000) 009; M. Carena, D. Garcia, U. Nierste and C. E. Wagner, hep-ph/0010003.
19. H. N. Brown *et al.* [Muon g-2 Collaboration], Phys. Rev. Lett. **86**, 2227 (2001) [hep-ex/0102017].
20. S. Mizuta and M. Yamaguchi, Phys. Lett. B **298** (1993) 120 [arXiv:hep-ph/9208251]; J. Edsjo and P. Gondolo, Phys. Rev. D **56** (1997) 1879 [arXiv:hep-ph/9704361]. C. Boehm, A. Djouadi and M. Drees, Phys. Rev. D **62** (2000) 035012 [arXiv:hep-ph/9911496]; J. Ellis, T. Falk and K. A. Olive, Phys. Lett. B **444**, 367 (1998); J. Ellis, T. Falk, K. A. Olive and M. Srednicki, Astropart. Phys. **13** (2000) 181; M. E. Gómez, G. Lazarides and C. Pallis, Phys. Rev. D **61**, 123512 (2000) [hep-ph/9907261] and Phys. Lett. B **487**, 313 (2000) [hep-ph/0004028]; R. Arnowitt, B. Dutta and Y. Santoso, hep-ph/0102181.
21. M. Drees and M. M. Nojiri, Phys. Rev. D **47**, 376 (1993); H. Baer and M. Brhlik, Phys. Rev. D **53** (1996) 597 and Phys. Rev. D **57** (1998) 567; H. Baer, M. Brhlik, M. A. Diaz, J. Ferrandis, P. Mercadante, P. Quintana and X. Tata, Phys. Rev. D **63** (2001) 015007; A. B. Lahanas, D. V. Nanopoulos and V. C. Spanos, hep-ph/0009065.

22. M. Battaglia *et al.*, arXiv:hep-ph/0106204.
23. J. Ellis, J. L. Feng, A. Ferstl, K. T. Matchev and K. A. Olive, arXiv:astro-ph/0110225.
24. J. Silk and M. Srednicki, Phys. Rev. Lett. **53** (1984) 624.
25. Y. Asaoka *et al.*, arXiv:astro-ph/0109007, and references therein.
26. J. Silk, K. A. Olive and M. Srednicki, Phys. Rev. Lett. **55** (1985) 257.
27. M. W. Goodman and E. Witten, Phys. Rev. D **31** (1985) 3059.
28. J. R. Ellis, R. A. Flores and S. Ritz, Phys. Lett. B **198** (1987) 393.
29. K. Greisen, Phys. Rev. Lett. **16** (1966) 748; G. T. Zatsepin and V. A. Kuzmin, JETP Lett. **4** (1966) 78.
30. A. Letessier-Selvon, astro-ph/0006111.
31. See, for example, P. G. Tinyakov and I. I. Tkachev, JETP Lett. **74** (2001) 1 [Pisma Zh. Eksp. Teor. Fiz. **74** (2001) 3] [arXiv:astro-ph/0102101].
32. P. G. Tinyakov and I. I. Tkachev, arXiv:astro-ph/0102476.
33. J. J. Beatty *et al.*, Auger Collaboration, *The Pierre Auger Project Design Report*, FERMILAB-PUB-96-024.
34. N. W. Evans, F. Ferrer and S. Sarkar, arXiv:astro-ph/0103085.
35. P. Blasi and R. Sheth, Phys. Lett. B **486** (2000) 233 [arXiv:astro-ph/0006316].
36. K. Benakli, J. Ellis and D. V. Nanopoulos, Phys. Rev. D **59** (1999) 047301.
37. J. Ellis, T. K. Gaisser and G. Steigman, Nucl. Phys. B **177** (1981) 427; J. Ellis, G. B. Gelmini, J. L. Lopez, D. V. Nanopoulos and S. Sarkar, Nucl. Phys. B **373** (1992) 399.
38. M. Birkel and S. Sarkar, Astropart. Phys. **9** (1998) 297.
39. D. J. Chung, E. W. Kolb and A. Riotto, Phys. Rev. D **59** (1999) 023501 and hep-ph/9810361; D. J. Chung, E. W. Kolb, A. Riotto and I. I. Tkachev, Phys. Rev. D **62** (2000) 043508.

40. O. Catalano, *Nuovo Cim.* **24C** (2001) 445.
41. For a review, see R. Mukherjee, [astro-ph/9901222](#).
42. R. Atkins *et al.*, Milagro Collaboration, [astro-ph/0001111](#).
43. G. Amelino-Camelia, J. Ellis, N. E. Mavromatos, D. V. Nanopoulos and S. Sarkar, *Nature* **393** (1998) 763; see also G. Amelino-Camelia, J. Ellis, N. E. Mavromatos and D. V. Nanopoulos, *Int. J. Mod. Phys. A* **12** (1997) 607;
44. P. Meszaros, [astro-ph/0111170](#) and references therein.
45. E. Waxman, [hep-ph/0004102](#) and references therein.
46. J. P. Norris, J. T. Bonnell, G. F. Marani and J. D. Scargle, [astro-ph/9912136](#).
47. J. Ellis, N. E. Mavromatos, D. V. Nanopoulos and G. Volkov, *Gen. Rel. Grav.* **32** (2000) 1777.
48. S. Hawking, *Commun. Math. Phys.* **87** (1982) 395; J. Ellis, J. S. Hagelin, D. V. Nanopoulos and M. Srednicki, *Nucl. Phys. B* **241** (1984) 381.
49. R. Gambini and J. Pullin, *Phys. Rev. D* **59** (1999) 124021.
50. H. Yu and L. H. Ford, [gr-qc/9907037](#) and [gr-qc/0004063](#); A. Campbell-Smith, J. Ellis, N. E. Mavromatos and D. V. Nanopoulos, *Phys. Lett. B* **466** (1999) 11.
51. L. Gonzalez-Mestres, [physics/0003080](#) and references therein; S. Coleman and S. L. Glashow, *Phys. Rev. D* **59** (1999) 116008 and references therein.
52. R. J. Protheroe and H. Meyer, [astro-ph/0005349](#).
53. F.A. Aharonian *et al.*, HEGRA Collaboration, *Astron. Astrophys.* **349** (1999) 11.
54. S. Barwick, for the AMANDA Collaboration, talk at Neutrino 2000.
55. J. R. Ellis, N. E. Mavromatos and D. V. Nanopoulos, *Phys. Rev. D* **63** (2001) 124025 [[arXiv:hep-th/0012216](#)].

QUANTUM-GRAVITY PHENOMENOLOGY WITH GAMMA RAYS AND UHE COSMIC RAYS

Giovanni Amelino-Camelia

Dipartimento di Fisica, Università "La Sapienza", P.le Moro 2, Roma, Italy

ABSTRACT

In recent years several ideas for experimental searches of effects induced by quantum properties of space-time have been discussed. Some of these ideas concern the role in quantum spacetime of the ordinary Lorentz symmetry of classical flat spacetime. Deviations from ordinary (classical) Lorentz symmetry are now believed to be rather natural in non-commutative spacetimes, models based on String Theory and models based on Loop Quantum Gravity. Observations of gamma rays and ultra-high-energy cosmic rays could play a key role in the development of this research programme.

1 Introduction

Quantum-Gravity Phenomenology ¹⁾ is an intentionally vague name for a new approach to research on the possible non-classical (quantum) properties of spacetime.

This approach does not adopt a specific formalism for the description of the short-distance structure of spacetime (*e.g.*, “string theory”, “loop quantum gravity” and “noncommutative geometry” are seen as equally deserving mathematical-physics programmes); it is rather the proposal that quantum-gravity research should proceed just in the familiar old-fashioned way: through small incremental steps starting from what we know, combining mathematical-physics studies with experimental studies to reach deeper and deeper layers of understanding of the short-distance structure of spacetime. For various “historical” reasons (mostly connected with the lack of guidance from experiments) research on quantum gravity has wandered off this traditional strategy: the most popular quantum-gravity approaches, such as string theory and loop quantum gravity, could be described as “top-to-bottom approaches”, since they start off with some key assumption about the structure of spacetime at the Planck scale and then they try (with limited, vanishingly small, success) to work their way back to “reality”, the realm of doable experiments. With “quantum-gravity phenomenology” I would like to refer to all studies that are somehow related with the “bottom-to-top approach”, consistently with traditional strategy of physics research.

Since the problem at hand is extremely difficult (arguably the most challenging problem ever faced by the physics community) it appears likely that the two complementary approaches might combine in a useful way: for the “bottom-to-top approach” it is important to get some guidance from the (however tentative) indications emerging from the “top-to-bottom approaches”, while for “top-to-bottom approaches” it might be very useful to be alerted by quantum-gravity phenomenologists with respect to the type of new effects that could be most effectively tested experimentally¹.

Until very recently the idea of a quantum-gravity phenomenology, and in particular of attempts of identification of experiments with promising sensitivity, was very far from the main interests of quantum-gravity research. One isolated idea had been circulating from the mid 1980s: it had been realized ^{2, 3, 4}) that the sensitivity of CPT tests using the neutral-kaon

¹It is hard for “top-to-bottom approaches” to obtain a complete description of low-energy physics, but perhaps it would be possible to dig out predictions on some specific spacetime features that appear to deserve special attention in light of the corresponding experimental sensitivities.

system is such that even small effects of CPT violation originating at the Planck scale² might in principle be revealed. These pioneering works on CPT tests were for more than a decade the only narrow context in which the implications of quantum gravity were being discussed in relation with experiments, but over the last 4 years several new ideas for tests of Planck-scale physics have appeared at increasingly fast pace, leading me to argue^{1, 5)} that the times might be right for a larger overall effort in this direction, which indeed could be called “quantum-gravity phenomenology”. At the present time (in addition to the already mentioned CPT tests) there are several examples of experimentally accessible contexts in which conjectured quantum-gravity effects are being considered, including studies of in-vacuo dispersion using gamma-ray astrophysics^{6, 7)}, studies of laser-interferometric limits on quantum-gravity induced distance fluctuations^{8, 9)}, studies of the role of the Planck length in the determination of the energy-momentum-conservation threshold conditions for certain particle-physics processes^{10, 11, 12, 13)}, and studies of the role of the Planck length in the determination of particle-decay amplitudes¹⁴⁾. These experimental/phenomenological studies might represent the cornerstones of quantum-gravity phenomenology since they are as close as one can get to direct tests of space-time properties, such as space-time symmetries. Other experimental proposals that should be seen as part of the quantum-gravity-phenomenology programme rely on the mediation of some dynamical theory in quantum space-time; comments on these other proposals can be found in Refs. ^{1, 15, 16, 17, 18, 19)}.

In these lecture notes I intend to emphasize those aspects of quantum-gravity-phenomenology that are relevant for the astrophysics community. The relevant topic is the one that concerns the faith of the Lorentz symmetry of classical spacetime when the spacetime is quantized. Since the Lorentz symmetry of classical flat (Minkowski) spacetime is verified experimentally to very high accuracy, it appears that any deviation from classical Lorentz symmetry, which might emerge from quantum-gravity

²The possibility of Planck-scale-induced violations of the CPT symmetry has been extensively considered in the literature. One simple point in support of this possibility comes from the fact that the CPT theorem, which holds in our present conventional theories, relies on exact locality, whereas in quantum gravity it appears plausible to assume lack of locality at Planckian scales.

theories, would be subject to severe experimental constraints. As a result Lorentz-symmetry tests are a key component of the programme of “quantum-gravity phenomenology” 1, 20, 21).

My main focus here will be on the faith of Lorentz invariance at the quantum-spacetime level. A large research effort has been devoted to this subject. Most of these studies focus on the possibility that Lorentz symmetry might be “broken” (in a sense clarified later in these notes) at the quantum level; however, I have recently shown that Lorentz invariance might be affected by spacetime quantization in a softer manner: there might be no net loss of symmetries but the structure of the Lorentz transformations might be affected by the quantization procedure^{22, 23}). In the following I shall describe rather pedagogically the main differences between the broken-symmetry and my new deformed-symmetry scenario. In addition I will comment on the type of astrophysical observations, involving gamma rays and ultra-high-energy cosmic rays, which could provide evidence of such symmetry-related quantum properties of space-time. An exciting recent development in this area is that certain puzzling gamma-ray and UHE cosmic-ray observations are being actively discussed as possible first manifestations of a quantum property of space-time.

Before going forward with these main points on my agenda for these lecture notes, let me make a parenthetical remark, further clarifying the objectives of quantum-gravity phenomenology: The primary challenge of quantum-gravity phenomenology is the one of establishing the properties of space-time at Planckian distance scales, since most theoretical arguments suggest that this is the characteristic scale of quantum space-time effects. However, there is also recent discussion of the possibility that quantum-spacetime effects might be stronger than usually expected, *i.e.* with a characteristic energy scale that is much smaller (perhaps in the TeV range!) than the Planck energy. Examples of mechanisms leading to this possibility are found in string-theory models with large extra dimensions²⁴) and in certain noncommutative-geometry models²⁵). Of course, the study of the phenomenology of these models is in the spirit of quantum-gravity phenomenology, but it is, in a sense, to be considered as a sideline development (and it is less challenging than the quantum-gravity-phenomenology efforts that pertain effects originating genuinely at the Planck scale).

2 The faith of Lorentz symmetry in quantum spacetime

If Nature hosts some form of “quantization” (even just in the general weak sense of “non-classical” properties) of space-time, this of course would also apply to flat spacetimes (*e.g.* if spacetime is in general discrete or noncommutative then of course the particular case of flat spacetime will also be described in the same way). One might argue (more or less convincingly) that quantum effects should be stronger in strong-curvature contexts, such as the ones involving black holes, but our capability of detailed experimental studies of such contexts is vanishingly small. Instead, in certain flat-spacetime contexts our experiments reach extremely high precision and therefore even relatively small effects induced by quantum properties of spacetime might be detectable. This is one of the key strategic points of my view on the development of quantum-gravity phenomenology ^{1, 5}).

In flat quantum spacetimes a key characteristic is the role of the Planck length, L_p . If the Planck length only has the role we presently attribute to it, which is basically the role of a coupling constant (an appropriately rescaled version of the gravitational coupling G), no problem arises for FitzGerald-Lorentz contraction, but if we try to promote L_p to the status of an intrinsic characteristic of space-time structure (or a characteristic of the kinematic rules that govern particle propagation in space-time) it is nearly automatic to find conflicts with FitzGerald-Lorentz contraction ^{22, 23}).

For example, it is very hard (perhaps even impossible) to construct discretized versions or non-commutative versions of Minkowski space-time which enjoy ordinary Lorentz symmetry. Pedagogical illustrative examples of this observation have been discussed, *e.g.*, in Ref. ²⁶) for the case of discretization and in Refs. ^{27, 28}) for the case of non-commutativity. The action of ordinary (classical) boosts on discretization length scales (or non-commutativity length scales) will naturally be such that different inertial observers would attribute different values to these lengths scales, just as one would expect from the mechanism of FitzGerald-Lorentz contraction.

There are also dynamical mechanisms (of the spontaneous symmetry-breaking type) that can lead to deviations from ordinary Lorentz invariance; it appears for example that this might be possible in String Field Theory ²⁹).

Both in String Theory and in Loop Quantum Gravity³ it is also natural to consider certain external-field backgrounds, which, in the appropriate sense ^{22, 23}) (they provide a way to identify a preferred class of inertial observers), break Lorentz invariance.

Departures from ordinary Lorentz invariance are therefore rather plausible at the quantum-gravity level. Here I want to emphasize that there are at least two possibilities: (i) Lorentz invariance is broken and (ii) Lorentz invariance is deformed.

2.1 Deformed Lorentz invariance

In order to be specific about the differences between deformed and broken Lorentz invariance let me focus on the dispersion relation $E(p)$ which will naturally be modified in either case. Let me also assume, for the moment, that the deformation be Planck-length induced: $E^2 = m^2 + p^2 + f(p, m; L_p)$. If the function f is nonvanishing and nontrivial and the energy-momentum

³As I shall argue more carefully elsewhere ³⁰), in Loop Quantum Gravity there might even be a fundamental departure from classical Lorentz invariance. This can be deduced from studies arguing that Loop Quantum Gravity predicts a fixed discrete spectrum of area eigenvalues, independently of the characteristic scale of curvature of the surface whose area is being measured (and therefore also for flat surfaces in flat spacetimes). One of the primary implications of Lorentz invariance is that the same experiment is seen by different observers in different ways which are however predictably (classically) connected by Lorentz transformations. If, for example, a series of measurements by one observer all give the same result of an area measurement, say the result A_0 , then according to classical Lorentz invariance those same measurements should be seen by another observer as measurements all giving the same but different, say A_1 , result (with A_1 related to A_0 by the appropriate boost). When the spectrum of the area of a flat surface in a flat spacetime is discrete this property of classical Lorentz invariance is at risk: the results A_0 being all the same would reflect the fact that one is dealing with what is an area eigenstate for observer O_0 , and A_0 should be an eigenvalue of the area operator, but, if the second observer O_1 is only minutely boosted with respect to O_0 , one should find that A_1 , the boosted value of A_0 , could not possibly be another eigenvalue (if the boost is small enough it will not be sufficient for reaching another eigenvalue in the discrete list of eigenvalues that composes the spectrum of the area operator) and it would be paradoxical for observer O_1 to find systematically repeated measurement results A_1 .

transformation rules are ordinary (the ordinary Lorentz transformations) then clearly f cannot have the exact same structure for all inertial observers. In this case one would speak of an instance in which Lorentz invariance is broken. If instead f does have the exact same structure for all inertial observers, then necessarily the transformations between these observers must be deformed. In this case one would speak of an instance in which the Lorentz transformations are deformed, but Lorentz invariance is preserved (in the deformed sense).

While much work has been devoted to the case in which Lorentz invariance is actually broken, the possibility that Lorentz invariance might be deformed was introduced only very recently by this author (22, 31, 32, 23, 33). An example in which all details of the deformed Lorentz symmetry have been worked out is the one in which one enforces as an observer-independent statement the dispersion relation

$$L_p^{-2} (e^{L_p E} + e^{-L_p E} - 2) - \vec{p}^2 e^{-L_p E} = m^2 \quad (1)$$

In leading (low-energy) order this takes the form

$$E^2 - \vec{p}^2 + L_p E \vec{p}^2 = m^2 . \quad (2)$$

The Lorentz transformations and the energy-momentum conservation rules are accordingly modified (22, 23, 33).

2.2 Broken Lorentz invariance

The case of broken Lorentz invariance requires fewer comments since it is more familiar to the community. In preparation for the following sections it is useful to emphasize that the same dispersion relation (2), which was shown in Refs. (22, 23) to be implementable as an observer-independent dispersion relation in a deformed-symmetry scenario, can also be considered ⁶⁾ as a characteristic dispersion relation of a broken-symmetry scenario. In this broken symmetry scenario the dispersion relation (2) would still be valid but only for one “preferred” class of inertial observers (*e.g.* the natural CMBR frame) and it would be valid approximately in all frames not highly boosted with respect to the preferred frame. In highly-boosted frames one might find the same form of the dispersion relation but with different value of the deformation scale (different from L_p). All this follows from the fact that in the broken-symmetry

scenario the laws of transformation between inertial observers are unmodified. Accordingly also energy-momentum conservation rules are unmodified.

Another scenario in which one finds broken Lorentz invariance is the one of canonical noncommutative spacetime, in which the dispersion relation is modified (with different deformation term ^{34, 35}), but, again, the energy-momentum Lorentz transformation rules are not modified. This example of noncommutative spacetime has been recently shown to be relevant for the description of string theory in certain external-field backgrounds (see, *e.g.*, Ref. 25, 34).

3 Illustrative example: photon-pair pion decay

Before discussing the role that observations of gamma rays and UHE cosmic rays could play in the development of this research area, let me clarify, in this Section, that the differences between the broken-symmetry and the deformed-symmetry case can be very significant for what concerns experimental signatures. This is also important since it proves that the relevant astrophysics observations might not only provide us the first manifestation of a quantum space-time property: they might even distinguish between different quantum pictures of spacetime.

In order to render very explicit the differences between the broken-symmetry and the deformed-symmetry case I consider here the simplest example in which these differences are rather dramatic: photon-pair pion decay. I adopt in one case deformed energy-momentum conservation ²³), as required by the deformed Lorentz transformations of the deformed-symmetry case, while in the other case I adopt ordinary energy-momentum conservation, as required by the fact that the Lorentz transformation rules are unmodified in the broken-symmetry case, but for both cases I impose the same dispersion relation (2).

In the broken-symmetry case, combining (2) with ordinary energy-momentum conservation rules, one can establish a relation between the energy E_π of the incoming pion, the opening angle θ between the outgoing photons and the energy E_γ of one of the photons (the energy of the second photon is of course not independent; it is given by the difference between the energy of the pion and the energy of the first photon):

$$\cos(\theta) = \frac{2E_\gamma E'_\gamma - m_\pi^2 + 3L_p E_\pi E_\gamma E'_\gamma}{2E_\gamma E'_\gamma + L_p E_\pi E_\gamma E'_\gamma}, \quad (3)$$

where indeed $E'_\gamma \equiv E_\pi - E_\gamma$. This relation shows that at high energies (starting at values of E_π of order $(m_\pi^2/L_p)^{1/3}$) the phase space available to the decay is anomalously reduced: for given value of E_π certain values of E_γ that would normally be accessible to the decay are no longer accessible (they would require $\cos\theta > 1$).

In the deformed-symmetry case one enforces the deformed conservation rules ²³⁾

$$E_\pi = E_\gamma + E'_\gamma, \quad \vec{p}_\pi = \vec{p}_\gamma + \vec{p}'_{\gamma'} + L_p E_\gamma \vec{p}'_{\gamma'}, \quad (4)$$

which, when combined again with (2), give rise to the different relation

$$\cos(\theta) = \frac{2E_\gamma E'_\gamma - m_\pi^2 + 3L_p E_\gamma^2 E'_\gamma + L_p E_\gamma E_\gamma'^2}{2E_\gamma E'_\gamma + 3L_p E_\gamma^2 E'_\gamma + L_p E_\gamma E_\gamma'^2}. \quad (5)$$

Here it is easy to check that one is never led to consider the paradoxical condition $\cos\theta > 1$. There is therefore no severe implication of the deformed-symmetry case for the amount of phase space available for the decays (certainly not at energies around $(m_\pi^2/L_p)^{1/3}$, possibly at Planckian energies).

4 An agenda for gamma-ray and UHECR studies

The key points for the phenomenology of quantum-gravity-induced deviations from classical Lorentz invariance are possible deformations of the dispersion relation and possible deformations of the energy-momentum conservation conditions.

Whether or not there is an accompanying deformation of energy-momentum conservation⁴ a deformation of the dispersion relation is expected to give rise to in vacuo dispersion ^{6, 1, 7)} and, possibly (if the space-time has corresponding structure ³⁵⁾), to birefringence. In vacuo dispersion would provide a striking signature: the speed of massless particles would depend on wavelength⁵ and therefore photons that we somehow know to have been emitted simultaneously up to $\Delta_0 T$ precision would reach us with relative time delays

⁴In the case of deformation of Lorentz symmetry both the dispersion relation and the energy-momentum conservation conditions are modified simultaneously, since they both must reflect ^{22, 23)} the structure of the deformed transformation rules between inertial observers.

⁵The ordinary dispersion relation is linear for massless particles, and

$\Delta_1 T$, where $\Delta_1 T > \Delta_0 T$, and one should also find some dependence of $\Delta_1 T$ on the amount of time the photon spent travelling in space-time (*i.e.* time spent under the influence of quantum properties of space-time). As discussed in Refs. 1, 6, 7) this type of effect can be naturally studied in the context of observations of gamma-ray bursts and observations of the high-energy photons emitted by certain blazars, such as Mk421. Certain gamma-ray observatories soon to be operational will have excellent sensitivity toward this type of effect, and in particular GLAST ³⁶⁾ is planning dedicated studies. Interest in such studies is also growing in AMS ³⁷⁾.

As discussed in the previous Section, also certain aspects of particle-decay physics, at high energies, may carry an important trace of quantum-space-time effects. In that context however the implications of a dispersion-relation deformation do depend strongly on whether there is an associated deformation of energy-momentum conservation (*i.e.* depend on whether one is dealing with a scenario with deformed symmetries or instead one is dealing with a scenario with broken symmetries). The outlook of these studies based on particle-decay anomalies is described in Ref. ¹⁴⁾, also using a related data analysis reported in Ref. ³⁸⁾.

But perhaps the most powerful tool for the experimental investigation of quantum-gravity-induced deviations from ordinary Lorentz invariance is provided by “threshold anomalies” ¹³⁾. It is to this topic, which deserves being discussed in detail, that I devote the remainder of the Section. It is intriguing to notice that the prediction of these threshold anomalies appears to be consistent with some puzzling results of astrophysics observations. In two different regimes, UHECRs and multi-TeV photons, the universe appears to be more transparent than expected. UHECRs should interact with the Cosmic Microwave Background Radiation (CMBR) and produce pions. TeV photons should interact with the Infra Red (IR) photons and produce electron-positron pairs. These interactions should make observations of UHECRs with $E > 5 \cdot 10^{19} \text{eV}$ (the GZK limit) ³⁹⁾ or of gamma-rays with $E > 10 \text{TeV}$ ⁴⁰⁾ from distant sources unlikely. Still UHECRs above the GZK limit and Mk501 photons with energies up to 24 TeV are observed.

therefore dE/dp is wavelength (energy) independent. A nonlinear Planck-length-deformed dispersion relation will instead inevitably lead to wavelength-dependent dE/dp .

A CMBR photon and a UHE proton with $E > 5 \cdot 10^{19}$ eV should satisfy the kinematic requirements (threshold) for pion production. UHE protons should therefore lose energy, due to photopion production, and should slow down until their energy is below the GZK energy. At higher energies the proton's mean free path decreases rapidly and it is down to a few Mpc at $3 \cdot 10^{20}$ eV. Yet more than 15 CRs have been observed with nominal energies at or above $10^{20} \pm 30\%$ eV ⁴¹⁾. There are no astrophysical sources capable of accelerating particles to such energies within a few tens of Mpc from us. Furthermore, if the CRs are produced homogeneously in space and time, we would expect a break in the CR spectrum around the GZK threshold, which is not seen.

HEGRA has detected high-energy photons with a spectrum ranging up to 24 TeV ⁴²⁾ from Markarian 501 (Mk501), a BL Lac object at a redshift of 0.034 (~ 157 Mpc). This observation indicates a second paradox of a similar nature. A high energy photon with energy E can interact with an IR background photon with wavelength $\lambda \sim 30 \mu m (E/10 \text{ TeV})$ and produce an electron-positron pair. The mean free path of TeV photons depends on the spectrum of the corresponding IR background. Recent data from DIRBE ^{43, 44, 45)} and from ISOCOM ⁴⁶⁾ suggest that the mean free path for 20 TeV photons should be much shorter than the one of 10 TeV photons. However, no apparent break is seen in the spectrum of Mk501 in the region 10-20 TeV.

The UHECR paradox is well established. Numerous theoretical models, mostly requiring new physics, have been proposed for its resolution (see Ref. ⁴⁷⁾ for a recent review). With much less data, and with some uncertainty on the IR background, the Mk501 TeV-photon paradox is less established. However, if indeed this must be considered as a paradox, there are no models for its resolution, apart from the possibility that the IR background estimates are too large. Planned experiments will soon provide us better data on both issues. At present it appears reasonable to assume, just as a working hypothesis, that both paradoxes are real.

The interpretation of these paradoxes as threshold anomalies is appealing for several reasons. In both paradoxes low-energy photons interact with high energy particles. The relevant reactions should take place at a kinematic threshold. In both cases the center-of-mass threshold energies are rather modest and the physical processes involved are well tested and understood. In spite of these similarities, so far, there is no model that explains both

paradoxes within a single theoretical scheme (unless the model accommodates an irritatingly large number of parameters). This appears to provide encouragement for the idea that quantum-gravity-induced deviations from ordinary Lorentz invariance might be responsible for both paradoxes⁶.

In order to illustrate the mechanism of threshold anomalies, let us consider, for example, the broken-symmetry case already considered in the preceding section. I will now apply it to the kinematics of the process of electron-positron pair production, which is relevant for the Mk501 paradox. Combining (2) with ordinary energy-momentum conservation rules, one can establish that at threshold the energy E of the hard Mk501 photon and the energy ϵ of the soft background photon must satisfy the relation $E \simeq m_e^2/\epsilon + L_p m_e^6/(8\epsilon^4)$. The correction $L_p m_e^6/(8\epsilon^4)$ is indeed sufficient to push the threshold energy upwards by a few TeV, consistently with the observations. As shown in Refs. ¹³⁾ (and references therein), an analogous result holds for the photopion threshold, which is relevant for the cosmic-ray paradox.

This type of analysis provides encouragement (of course, very preliminary) for the hypothesis that the two paradoxes might be the first ever manifestation of a quantum (Planck-length related) property of spacetime.

Just like in the case of pion decay, considered in the preceding Section, also for the evaluation of threshold anomalies there are large quantitative differences (which will be discussed in detail in a paper now in preparation ⁵⁰⁾) between the case in which Lorentz symmetry is broken and the case in which Lorentz symmetry is deformed. More accurate information on the paradoxes, such as the one that will be provided by Auger ⁵¹⁾, can therefore even start pointing us toward the proper language for the description of the short-distance (quantum) structure of spacetime.

Experimental studies such as the ones planned by Auger will also in general clarify whether the origin of the paradoxes is indeed kinematical. I want to stress that, in this respect, it is important to get high-quality data

⁶It is of course also possible to consider deviations from Lorentz invariance that do not have quantum-gravity origin ^{48, 49)}, but, as discussed in Refs. ¹³⁾, the idea of a quantum-gravity origin, besides being conceptually appealing, leads to a natural estimate for the magnitude of the effects, and this estimate appears to fit well the observations (while non-quantum-gravity approaches host a large number of parameters to be freely adjusted to obtain the needed magnitude of the departure from Lorentz-invariance).

in the neighbourhood of the expected GZK cutoff, perhaps even more than establishing how far (how high in energy) the cosmic-ray flux extends. In fact, the kinematical mechanism of threshold anomalies leads to the definite general prediction that nothing at all particular should happen at the GZK scale, since the GZK threshold is simply moved forward (or eliminated all together ^{13, 52}) by the deviations from classical Lorentz invariance. Other attempts of explaining the cosmic-ray paradox instead must coexist with the GZK threshold and therefore (unless huge parametric fine-tuning is allowed) will inevitably predict at least some peculiarity to occur at the GZK scale.

References

1. G. Amelino-Camelia, "Are we at the dawn of quantum-gravity phenomenology?", *Lect. Notes Phys.* **541**, 1 (2000).
2. J. Ellis, J.S. Hagelin, D.V. Nanopoulos and M. Srednicki, *Nucl. Phys. B* **241**, 381 (1984).
3. P. Huet and M.E. Peskin, *Nucl. Phys. B* **434**, 3 (1995).
4. V.A. Kostelecky and R. Potting, *Phys.Rev. D* **51**, 3923 (1995); O. Bertolami, D. Colladay, V.A. Kostelecky and R. Potting, *Phys.Lett. B* **395**, 178 (1997).
5. G. Amelino-Camelia, gr-qc/0012049, *Nature* **408**, 661 (2000).
6. G. Amelino-Camelia, J. Ellis, N.E. Mavromatos, D.V. Nanopoulos and S. Sarkar, astro-ph/9712103, *Nature* **393**, 763 (1998).
7. S.D. Biller *et al.*, *Phys. Rev. Lett.* **83**, 2108 (1999).
8. G. Amelino-Camelia, gr-qc/9808029, *Nature* **398**, 216 (1999).
9. G. Amelino-Camelia, gr-qc/0104005.
10. T. Kifune, *Astrophys. J. Lett.* **518**, L21 (1999).
11. W. Kluźniak, *Astropart. Phys.* **11**, 117 (1999).
12. R.J. Protheroe and H. Meyer, *Phys. Lett. B* **493**, 1 (2000).

13. G. Amelino-Camelia and T. Piran, *Phys. Lett. B* **497**, 265 (2001);
Phys. Rev. D **64**, 036005 (2001).
14. G. Amelino-Camelia, gr-qc/0107086.
15. R. Brustein, gr-qc/9810063; G. Veneziano, hep-th/9902097; M. Gasperini,
hep-th/9907067.
16. I.C. Perival and W.T. Strunz, *Proc. R. Soc. A* **453**, 431 (1997).
17. L.J. Garay, *Phys. Rev. Lett.* **80**, 2508 (1998).
18. D.V. Ahluwalia, gr-qc/9903074, *Nature* **398**, 199 (1999).
19. C. Lämmerzahl and C.J. Bordé, *Lect. Notes Phys.* **562**, 463 (2001).
20. C. Rovelli, *Notes for a brief history of Quantum Gravity*, gr-qc/0006061,
in "Proceedings of the 9th Marcel Grossmann Meeting on Recent
Developments in Theoretical and Experimental General Relativity,
Gravitation and Relativistic Field Theories, Rome, Italy, 2-9 Jul 2000".
21. S. Carlip, *Rept. Prog. Phys.* **64**, 885 (2001).
22. G. Amelino-Camelia, gr-qc/0012051, *Int. J. Mod. Phys. D* (2001) in press;
hep-th/0012238, *Phys. Lett. B* **510**, 255 (2001).
23. G. Amelino-Camelia, gr-qc/0106004.
24. N. Arkani-Hamed, S. Dimopoulos and G. Dvali, *Phys. Lett. B* **429**, 263
(1998).
25. M. R. Douglas and N. A. Nekrasov, hep-th/0106048.
26. G. 't Hooft, *Class. Quant. Grav.* **13**, 1023 (1996).
27. S. Majid and H. Ruegg, *Phys. Lett. B* **334**, 348 (1994).
28. J. Lukierski, H. Ruegg and W.J. Zakrzewski *Ann. Phys.* **243**, 90 (1995).
29. V.A. Kostelecky and R. Potting, *Phys.Lett. B* **381**, 89 (1996).
30. G. Amelino-Camelia, in preparation.
31. J. Kowalski-Glikman, hep-th/0102098.

32. G. Amelino-Camelia and M. Arzano, hep-th/0105120.
33. N.R. Bruno, G. Amelino-Camelia and J. Kowalski-Glikman, *Phys.Lett. B* **522**, 133 (2001).
34. A. Matusis, L. Susskind and N. Toumbas, *JHEP* **0012**, 002 (2000).
35. G. Amelino-Camelia, L. Doplicher, S. Nam and Y.-S. Seo, hep-th/0109191.
36. J.P. Norris, J.T. Bonnell, G.F. Marani, J.D. Scargle, astro-ph/9912136; A. de Angelis, astro-ph/0009271. (Additional information on GLAST is available at <http://glast.gsfc.nasa.gov/>)
37. Private communication, R. Battiston, AMS collaboration (updated information on these particular plans of investigation, including slides from an invited talk given to the AMS collaboration by this author, can be found at <http://ams.cern.ch/AMS/Analysis/gamma-pos/gamma-pos.html>).
38. E.E. Antonov, L.G. Dedenko, A.A. Kirillov, T.M. Roganova, G.F. Fedorova and E.Yu. Fedunin, *JETP Lett.* **73**, 446 (2001).
39. K. Greisen, *Phys. Rev. Lett.* **16** 748 (1966); G.T. Zatsepin and V.A. Kuzmin., *Sov. Phys.-JETP Lett.* **4**, 78 (1966).
40. A.I. Nikishov, *Sov. Phys. JETP* **14**, 393 (1962); R.J. Gould and G.P. Schreder, *PR* **155**, 1404 (1967); F.W. Stecker, O.C. De Jager and M.H. Salomon, *ApJ* **390L**, L49 (1992).
41. M. Takeda *et al.*, *Phys. Rev. Lett.* **81**, 1163 (1998).
42. F.A. Aharonian *et al.*, *A&A* **349**, 11A (1999).
43. E.L. Wright, astro-ph/0004192.
44. D.P. Finkbeiner, M. Davis and D.J. Schlegel, astro-ph/0004175.
45. M.G. Hauser *et al.*, *ApJ* **508**, 25 (1998).
46. A. Biviano, *et al.*, astro-ph/9910314 (in the proceedings of the 1999 Marseille IGRAP conference “Clustering at High Redshifts”).
47. A.V. Olinto, *Phys. Rept.* **333-334**, 329 (2000).

48. L. Gonzalez-Mestres, physics/9704017.
49. S. Coleman, S.L. Glashow, Phys. Rev. **D59**, 116008 (1999).
50. G. Amelino-Camelia, in preparation.
51. Information on the Pierre Auger Observatory is available at <http://www.auger.org/>.
52. Y.J. Ng, D.S. Lee, M.C. Oh and H. van Dam, Phys. Lett. **B507**, 236 (2001).

HIGH PRECISION COSMOLOGY

Amedeo Balbi

Dipartimento di Fisica - Università di Roma 'Tor Vergata'

INFN - Sezione di Roma II

Via della Ricerca Scientifica 1, I-00133 Roma, Italy

ABSTRACT

I review the current status of cosmology as emerging from recent observations of cosmic microwave background anisotropies as well as from other sources of cosmological information.

1 Introduction

The widely accepted paradigm for cosmology is the hot Big Bang model. In this framework, the geometry and evolution of the Universe is defined by its matter and energy content through general relativity theory. The Universe is expanding, so that it was hotter and denser at earlier times. The rate of expansion is quantified by the Hubble parameter H , whose present value H_0 is parameterized by the quantity h as $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$. The amount of matter and energy in the Universe from different components (baryons, dark

matter, radiation, vacuum energy, etc.) is parameterized by the quantities $\Omega_{(i)} \equiv \rho_{(i)}/\rho_c$. The critical density, $\rho_c = 1.88 \times 10^{-29} h^2 \text{ g cm}^{-3}$, is defined in such a way that $\Omega \equiv \sum_i \Omega_{(i)} = 1$ for a Universe with flat geometry (while $\Omega < 1$ and $\Omega > 1$ for open and closed geometry respectively).

An additional ingredient of the standard cosmological model is *inflation*¹⁾, a phase of early superluminal expansion of the Universe required to solve some problems of the Big Bang model. Inflation makes some well-defined predictions. First of all, the geometry of the Universe has to be very close to flat. Second, the structure we observe today in the Universe was produced by gravitational amplification of primordial density perturbations generated during inflation, characterized by having a nearly scale-invariant spectrum and by being Gaussian distributed.

While until recent times the knowledge of the parameters of the cosmological model was plagued by large uncertainties, the situation has now dramatically changed. Cosmology is not a data-starved science anymore. In the past few years, high-quality observations have fueled an impressive progress in our understanding of the Universe. We have entered the epoch of high precision cosmology.

Recent results from observation of the CMB temperature anisotropy have allowed us to constrain most cosmological parameters to unprecedented accuracy, giving for the first time a robust determination of the total energy density (and in turn of the geometry) of the Universe. In addition, a whole set of new observations of the large-scale structure properties of the Universe have put the determination of the mean matter density in the Universe on a firm ground. Finally, measurements of distant Type Ia Supernovae have recently provided evidence that the Universe has just entered a phase of accelerated expansion. In the following I will review the emerging scenario, giving particular emphasis to CMB as a cosmological probe.

2 Cosmology with the Cosmic Microwave Background

The Cosmic Microwave Background (CMB) is a snapshot of the infant Universe, when it was just about 300 000 years old. According to the standard Big Bang model, before that epoch the temperature in the Universe was so high that no neutral atom could stably exist. The Universe was basically a plasma of mainly free electrons and protons, kept in equilibrium with photons

by frequent Thomson scattering. Later, the Universe cooled down as a result of the expansion, and neutral atoms began to form. The photons could then decouple from the matter and travel freely, being finally observed today as an almost uniform background. The fact that the CMB was indeed found to have a black-body spectrum (a clear signature of the early period of matter-radiation equilibrium) with an astonishing precision ²⁾ is one of the big successes of the Big Bang model.

Since the distribution of the CMB photons reflects that of matter at the time of decoupling, any inhomogeneities in the matter density (needed to seed structure formation by gravitational instability) must leave an imprint as fluctuations of the CMB temperature. The presence of these *CMB temperature anisotropies* was first detected by NASA's COBE satellite in the early 90's ³⁾. The fact that the level of anisotropy is very small (about a part in one thousand, corresponding to temperature fluctuations of some tens of μK) simplifies the task of making theoretical prediction of the anisotropy pattern, since linear perturbation theory can be applied.

The bulk of the cosmological information encoded in the anisotropy pattern is concentrated at angular scales smaller than about 1 degree on the sky, corresponding to perturbations that were inside the horizon (i.e. in causal contact) before decoupling. On these scales, physical processes in the early Universe were able to leave their imprint on the CMB. For this reason, over the last decade a large number of ground-based and balloon-borne experiments performed observations of the fine-structure pattern of the anisotropy.

The observed temperature fluctuation in a given direction of the sky can be expanded in spherical harmonics:

$$\frac{\Delta T}{T}(\theta, \phi) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi). \quad (1)$$

The coefficients $C_l \equiv \langle |a_{lm}|^2 \rangle$ define the *angular power spectrum* of the CMB anisotropy¹. Because the Universe is isotropic on average, the C_l 's do not depend on the azimuthal index m . If the primordial density fluctuations are

¹The symbol $\langle \cdot \rangle$ represents an average over the statistical ensemble. Since we can only observe one realization of the ensemble — our own sky — we can at best build an un-biased estimate of C_l from the observations. This is: $C_l \equiv \frac{1}{(2l+1)} \sum_{m=-l}^l |a_{lm}|^2$.

Gaussian distributed, the angular power spectrum C_l fully characterizes the statistics of the temperature anisotropy pattern. The power spectrum is then the main CMB observable. Since each l is related to an angular scale θ on the sky given approximately by $l \sim \pi/\theta$, the power spectrum at high l 's probes sub-horizon angular scales at the time of decoupling and carries the imprint of physical processes which occurred in the early Universe. Conversely, low l 's probe the primordial shape of the power spectrum².

The way the shape of the CMB angular power spectrum depends on cosmology can be understood by simple physical considerations. Let us consider a density fluctuation of given physical scale in the baryon-photon fluid. Let us suppose that the physical scale of the fluctuation is smaller than the horizon size at decoupling, so that the inner region of the fluctuation is in causal contact. The amplitude of perturbation in the baryon component tends to be amplified by gravitational collapse. However, the radiation pressure provided by the photons prevents the collapse from happening. These competing mechanisms sets up harmonic oscillations in the amplitude of the perturbation. Since the amount of resistance to compression is quantified by the sound velocity in the fluid, this oscillations are called *acoustic*. When the photons decouple from matter, perturbations having different physical scale are caught in a different stage of oscillation and then have a different amplitude. The CMB photons we receive today carry this phase information as fluctuations in their temperature at different angular scales. This reflects in a series of harmonic acoustic peaks in the CMB angular power spectrum.

For a given initial distribution of density perturbations in the early Universe, the height of the acoustic peaks is mostly affected by the amount of matter in the Universe. If we enhance the baryon content of the Universe, keeping fixed all the other components, the compression stage of the fluid is more effective, increasing the amplitude of fluctuations at decoupling. Then, the relative height of the peaks in the CMB power spectrum represents a good indicator of the density of baryonic matter in the Universe. On the other hand, the position of the peaks depends on the way a certain physical scale at decoupling is mapped into an angular dimension on the sky. This is quantified, in a given cosmological model, by the so called *angular diameter distance relation*.

²Of course, neglecting secondary processes which may alter the CMB photon distribution after decoupling.

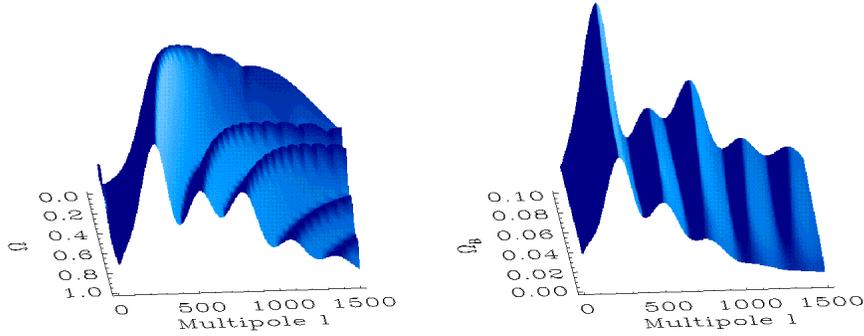


Figure 1: *The effect of cosmological parameters on the peak structure of the CMB angular power spectrum. On the left, the effect of varying the total energy density while keeping all the other parameters fixed. On the right, the effect of varying the baryon density.*

This relation mainly depends on the geometry of the Universe: in an open Universe, a certain physical scale at decoupling is seen today under a smaller angle than in a flat Universe. So, the position of the peaks in the CMB angular power spectrum is a good indicator of the geometrical properties of the Universe. The dependence of the CMB angular power spectrum on the geometry of the Universe and on the baryon density is shown in Figure 1.

3 Constraints on Cosmological Parameters from the CMB

The quality of CMB observations has considerably improved in recent times. The balloon-borne observations carried on by the BOOMERanG ⁴⁾ and MAXIMA ⁵⁾ teams (from Antarctica and from Texas, respectively) have produced the first images of the fine-scale pattern of CMB temperature anisotropy. The CMB map from BOOMERanG covers a 1800 square degrees patch of the southern sky. MAXIMA mapped a 124 square degrees patch of the northern sky. More recently, the DASI ⁶⁾ team released new maps over 32 sky fields of 3.4 degrees in diameter, obtained using ground-based interferometry from Antarctica. The kind of spatial features observed by these three independent experiments in different sky regions looks quite similar (see Figure 2).

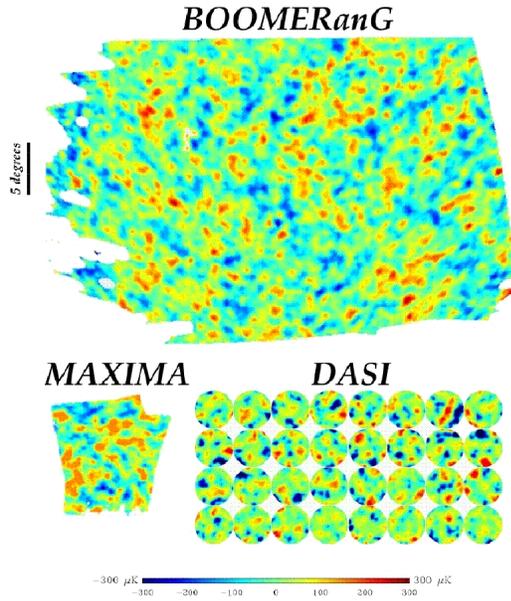


Figure 2: *Maps of the CMB temperature anisotropy produced by the BOOMERanG, MAXIMA and DASI experiments.*

From these observations, estimates of the CMB angular power spectrum have been obtained over a large range of multipoles ($20 \leq l \leq 1200$; see Figure 3). The power spectra measured by BOOMERanG, MAXIMA and DASI are in remarkable agreement and show unambiguously the presence of a sharp peak in the region $180 \leq l \leq 220$, as well as evidence of excess power at higher l 's, consistent with the presence of a second and third peak.

Likelihood analyses of these power spectrum measurements have been performed by each team to set constraints on the value of cosmological parameters. They agree about the fact that the CMB data strongly favor a Universe with flat geometry, and with scale-invariant primordial density fluctuations: the inflationary scenario brilliantly passed two important tests. Furthermore, the baryon density derived from the CMB is in striking agreement with the value resulting from comparing the measured primordial light elements abundances

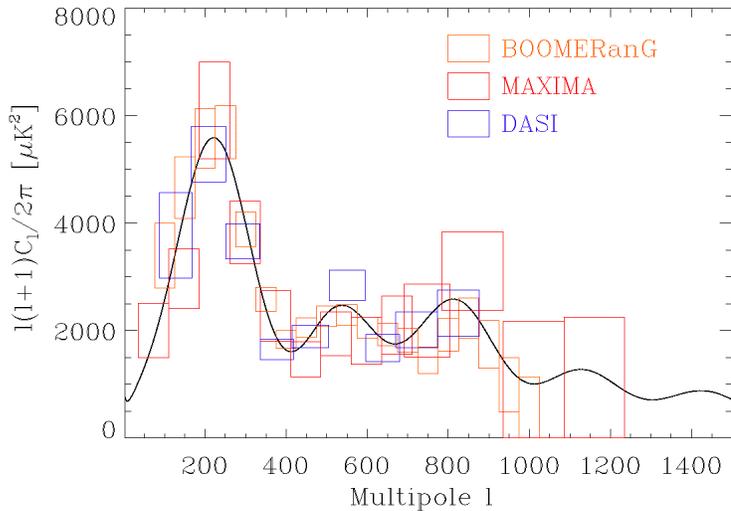


Figure 3: *Measurements of the CMB angular power spectrum from BOOMERanG, MAXIMA and DASI. The continuous line a reference theoretical model for a flat cosmology.*

with the big bang nucleosynthesis (BBN) predictions: $\Omega_b h^2 = 0.020 \pm 0.002$ ⁷⁾. This is an important indication of the self-consistency of our cosmological model, since the CMB and BBN values for the baryon density are obtained using entirely different methodology and observations.

4 The Concordance Model

The success of the CMB in giving us a reliable estimate of the total energy density of the Universe leaves us with the problem of finding out which is the contribution from different components to the critical density. Measuring the mean mass density of the Universe with traditional cosmological observations has always been a difficult task. Large enough samples have to be observed in order to be representative of the whole Universe. Furthermore, the distribution of matter cannot be directly deduced from that of light. However, the matter density is currently constrained by a number of independent and consistent ob-

servations (baryon-to-total mass ratio in clusters of galaxies, peculiar velocities and bulk flows, redshift surveys) to be roughly 1/3 of the total energy density ($\Omega_M = 0.33 \pm 0.04$ ⁸⁾). Where does the rest of critical density comes from?

Observations of distant type Ia supernovae ⁹⁾ recently allowed to probe the classic Hubble diagram up to very high redshifts. The surprising result was that, contrarily to expectations, the Universe is speeding up rather than slowing down. The fact that we are now entering a phase of cosmic acceleration has been explained with the presence of a smooth, negative-pressure component, which has been named *dark energy*. The best candidate for dark energy is a cosmological constant, or vacuum energy, i.e. the vacuum expectation value of some fundamental scalar field.

Cosmological models with flat geometry but different amount of vacuum energy have almost the same angular diameter distance relation. This makes the CMB angular power spectrum basically unable to distinguish which fraction of the critical density is provided by matter and which by the vacuum energy. However, when we look at the constraints in the Ω_M — Ω_Λ plane coming from the CMB, the observation of large-scale structure (LSS) and type Ia supernovae (SN Ia) an interesting picture emerges (see Figure 4). The CMB and the LSS suggest that 2/3 of the critical density must be provided by vacuum energy. The CMB and the SN Ia get to the same conclusion. The three constraints taken together identify a *concordance* region in the parameter space where $\Omega_M \sim 1/3$, $\Omega_\Lambda \sim 2/3$, and $\Omega = \Omega_M + \Omega_\Lambda \sim 1$. The fact that three independent and different kinds of observation, each probing a different epoch of the cosmic evolution and different physical processes, have converged to give a coherent picture is a big success of cosmology.

5 Future Prospects

While a consistent and reliable picture of the Universe is emerging, there are still open questions. One of the most puzzling aspects is the nature of the dark energy which seems to be the main contribution to the density of the Universe. The vacuum energy estimated from quantum field theory (as vacuum expectation value of some fundamental quantum field) is 10^{122} to 10^{55} times larger than the observed one, which leads to an extreme *fine-tuning* problem. Furthermore, vacuum energy is dominating the cosmic expansion right now, which seems to make the present epoch a very special one in the evolution of the

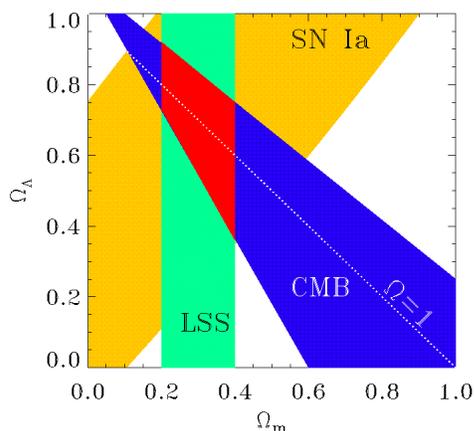


Figure 4: Likelihood contours (95% confidence level) from CMB, supernovae and large-scale structure observations.

Universe (*coincidence problem*). These problems are mitigated in the so-called *quintessence* models, where the scalar field responsible for the vacuum energy contribution is evolving through an equation that admits *tracking solutions*: large set of initial conditions result in the same vacuum energy at present. Attempts to use current CMB data to investigate the nature of dark energy have recently been made ¹⁰).

Future CMB missions from space will shed more light on this and other open problems. The NASA's MAP mission³ is currently operating and will soon produce full sky maps of the CMB sky at high angular resolution. In 2007 the ESA's Planck satellite⁴ will measure CMB temperature and polarization over the full sky with unprecedented angular resolution and instrumental sensitivity, reaching the theoretical limit in the power spectrum measurement over a large range of multipoles ($2 \leq l \leq 3000$). This observations, together with other sources of information (most notably further supernovae measurements from space such as those expected from the SNAP satellite⁵ and redshift surveys

³<http://map.gsfc.nasa.gov/>

⁴<http://astro.estec.esa.nl/Planck>

⁵<http://snap.lbl.gov>

such as SDSS⁶) will further strengthen our understanding of the Universe.

6 Acknowledgements

It is a pleasure to thank the organizers for the invitation and for the enjoyable and stimulating environment.

References

1. Guth, A.H., *Phys Rev* **D23**, 347 (1981); for a recent pedagogical exposition see Albrecht, A., astro-ph/0007247 (2000).
2. Fixsen, D.J., et al., *ApJ*, **473**, 576 (1996).
3. Smoot, G.F., et al., *ApJ*, **396**, L1 (1992).
4. de Bernardis, P., et al., *Nature*, **404**, 955 (2000); Lange, A.E., et al., *Phys. Rev.* **D63** 042001 (2001); de Bernardis, P., et al., *ApJ*, in press, astro-ph/0105296 (2001); Netterfield, C.B., et al., *ApJ*, in press, astro-ph/0104460 (2001).
5. Balbi, A., et al., *ApJL*, **545**, L1 (2000); Hanany, S., et al., *ApJL*, **545**, L5 (2000); Lee, A.T., et al., *ApJL*, in press, astro-ph/0104459 (2001); Stompor, R., et al., *ApJL*, in press, astro-ph/0105062 (2001).
6. Halverson, N.W., et al., *ApJ*, in press, astro-ph/0104489 (2001); Pryke, C., et al., *ApJ*, in press, astro-ph/0104490 (2001).
7. Burles, S., Nollett, K.M., Truran, J. N., & Turner, M. S. *Phys. Rev. Lett.*, **82**, 4176 (1999).
8. Turner, M.S. submitted to *ApJ*, astro-ph/0106035 (2001).
9. Perlmutter, S., et al., *ApJ*, **517**, 565 (1999); Riess, A.G., et al., *AJ*, **116**, 1009 (1998).
10. Balbi, A., Baccigalupi, C., Matarrese, S., Perrotta, F. & Vittorio, N., *ApJ*, **547**, L89 (2000); Baccigalupi, C., Balbi, A., Perrotta, F., Matarrese, S. & Vittorio, N., submitted to *Phys. Rev. D*, astro-ph/0109097 (2001)

⁶<http://www.sdss.org>

Cosmic Rays

E. Berezhko	Cosmic Ray Acceleration and High-Energy Gamma-Ray Production in Supernova Remnants
M.Simon	The interpretation of radioactive isotopes in the cosmic radiation and the link between the diffusion halo model and the leaky box model
F.Donato	Propagation of galactic cosmic ray nuclei and antiprotons in a diffusion model
P.Spillantini	Balloon experiments
M.Boezio	Analysis of cosmic-ray data. Examples with antiprotons and positrons
Y. Butt	Supernova Remnants as Hadronic Cosmic Ray Accelerators?
M.Casolino	Radiation in Space: Causes Measures and Countermeasures.
G.Esposito	Protons and Lepton Trapped in the Earth's Radiation Belts
R.Sparvoli	Cosmic rays in the Earth vicinity
P.Spillantini	The antimatter component of cosmic rays and the PAMELA experiment
R.Battiston	The ALPHA Magnetic Spectrometer, a Particle Physics Experiment in Space
L.Scarsi	EUSO: Using High Energy Cosmic Rays and Neutrinos as Messengers from the Unknow Universe
C. Sbarra	Photomultiplers for the New TOF System of AMS-02 Space Experiment

COSMIC RAY ACCELERATION AND HIGH-ENERGY GAMMA-RAY PRODUCTION IN SUPERNOVA REMNANTS

E.G. Berezhko
*Institute of Cosmophysical Research and Aeronomy,
Lenin Ave. 31, 677891 Yakutsk, Russia*

ABSTRACT

The purpose of this talk is to review some recent work on the model for the origin of the bulk of the galactic cosmic rays (CRs), namely that they are produced by diffusive shock acceleration in shock waves associated with supernova remnants (SNRs). This is currently the modern theory for the origin of galactic CRs. Selfconsistent nonlinear theory of CR acceleration in SNRs, developed during the last decade, is able to explain the main characteristics of the observed CR spectrum under several reasonable assumptions, at least up to an energy of 10^{15} eV. A brief review of the experimental results for searching of high energy gamma-ray emission from nearby SNRs and their correspondence to the theoretical predictions is presented.

1 Introduction

Considerable efforts have been made during the last years to empirically confirm the theoretical expectation that the main part of the Galactic cosmic rays (CRs) originates

in supernova remnants (SNRs). Theoretically progress in the solution of this problem has been due to the development of the theory of diffusive shock acceleration (see, for example, reviews ^{1, 2, 3}). Although still incomplete, the theory is able to explain the main characteristics of the observed CR spectrum under several reasonable assumptions, at least up to an energy of $10^{14} \div 10^{15}$ eV. Direct information about the dominant nucleonic CR component in SNRs can only be obtained from γ -ray observations. If this nuclear component is strongly enhanced inside SNRs then through inelastic nuclear collisions, leading to pion production and subsequent decay, γ -rays will be produced.

CR acceleration in SNRs ^{4, 5, 6, 7}, and the properties of the associated γ -ray emission ^{8, 9} were investigated in a number of studies (we mention here only those papers which include the effects of shock geometry and time-dependent nonlinear CR backreaction; for a review of others which deal with the test particle approximation, see for example ^{1, 2, 3}). All of these studies are based on a two-fluid hydrodynamical approach and directly employ the assumption that the expanding SN shock is locally plane; as dynamic variables for the CRs the pressure and the energy density are determined. Their characteristics are sometimes essentially different from the results obtained in a kinetic approach ^{10, 11, 12} which consistently takes the role of shock geometry and nonlinear CR backreaction into account. First of all, in kinetic theory the form of the spectrum of accelerated CRs and their maximum energy are calculated selfconsistently. In particular, the maximum particle energy ϵ_{max} , achieved at any given evolutionary stage, is determined by geometrical factors ¹³, in contrast to the hydrodynamic models which in fact postulate that the value of $\epsilon_{max}(t)$ is determined by the time interval t that has passed since the explosion ^{4, 5, 6, 7}. Although the difference between the values of ϵ_{max} in the two cases is not very large, it critically influences the structure and evolution of the shock. For example, the shock never becomes completely modified (smoothed) by the CR backreaction ^{10, 14}. Together with the smooth precursor, the shock transition always contains a relatively strong subshock which heats the swept-up gas and leads to the injection of suprathermal gas particles into the acceleration process. This prediction is in agreement with the observations that show significant gas heating in young SNRs.

A brief review of the kinetic model of CRs acceleration and subsequent γ -ray production inside SNRs is presented below.

2 Kinetic model

During the early phase of SNR evolution the hydrodynamical SN explosion energy E_{sn} is kinetic energy of the expanding shell of ejected mass. The motion of these ejecta produces a strong shock wave in the background interstellar medium (ISM), whose size R_s increases with velocity $V_s = dR_s/dt$. Diffusive propagation of energetic particles in the collisionless scattering medium allows them to traverse the shock front many times. Each two subsequent shock crossings increase the particle energy. In plane geometry this diffusive shock acceleration process creates a power law-type CR momentum spectrum. Due to their large energy content the CRs can dynamically modify the shock structure.

The description of CR acceleration by a spherical SNR shock wave is based on the diffusive transport equation for the CR distribution function. The gas matter is described by the gas dynamic equations which include the CR backreaction via term $-\partial P_c/\partial r$, i.e. the gradient of CR pressure. They also describe the gas heating due to the dissipation of Alfvén waves in the upstream region.

The SNR shock always includes a sufficiently strong subshock which heats the gas and plays an important dynamical role. The gas subshock, situated at $r = R_s$, is treated as a discontinuity on which all hydrodynamical quantities undergo a jump. The injection of some (small) fraction of gas particles into the acceleration process takes place at the subshock.

At present we only have some experimental¹⁵⁾ and theoretical^{16, 17)} indications as to what value of the injection rate can be expected. A simple CR injection model, in which a small fraction η of the incoming particles is instantly injected at the gas subshock with a speed $\lambda > 1$ times the postshock gas sound speed c_{s2} , is usually used^{10, 11, 12)}.

It is usually assumed that the Bohm diffusion coefficient $\kappa(p) = \rho_B c/3$ is a good approximation for strong shocks, characterized by intense Alfvén wave generation by accelerated CRs. Here ρ_B is the gyroradius of a particle with momentum p in the magnetic field B , c is the speed of light.

Alfvén wave dissipation as an additional heating mechanism strongly influences the structure of a modified shock in the case of large sonic Mach number $M = V_s/c_s \gg \sqrt{M_a}$, $M_a = V_s/c_a$ is the Alfvénic Mach number, c_s and c_a are the local sound and Alfvén speeds correspondingly, at the shock front position $r = R_s$. The wave damping substantially restricts the growth of the shock compression ratio $\sigma = \rho_2/\rho_s$ at the level $\sigma \approx M_a^{3/8}$ which, in the absence of Alfvén wave dissipa-

tion, has been found to reach extremely high values $\sigma \approx M^{3/4}$ for large Mach numbers (10, 18).

3 CR spectrum and composition

The main fraction of the galactic volume is occupied by so-called hot and warm phases of ISM, with hydrogen number density, temperature and magnetic field values $N_H = 0.003 \text{ cm}^{-3}$, $T_0 = 10^6 \text{ K}$, $B_0 = 3 \mu\text{G}$ and $N_H = 0.3 \text{ cm}^{-3}$, $T_0 = 10^4 \text{ K}$, $B_0 = 5 \mu\text{G}$ respectively. The ISM temperature T_0 determines the equilibrium ionization state of elements: at $T_0 = 10^4 \text{ K}$ Q_0 is close to 1 for all elements, whereas at $T_0 = 10^6 \text{ K}$ mean ion charge number increases from $Q_0 \approx 1$ for H and He to $Q_0 \approx 10$ for heavy ions with $A \approx 100$.

Expected CR spectra produced in SNRs calculated¹⁹⁾ at the SN explosion energy $E_{sn} = 10^{51} \text{ erg}$ and the ejecta mass $M_{ej} = 1.4M_\odot$, which are typical for SN Ia expanding into the uniform ISM, moderate injection rate $\eta = 10^{-4}$ are compared in Fig.1 with the experimental data. The softening of CR spectra in the Galactic confinement volume due to the rigidity dependent mean CR residence time $\tau(R) \propto R^{-\alpha}$ and the solar wind modulation effect are taken into account. The value of α and the proton spectrum normalization are selected to fit the experiment. The all particle spectrum $J_\Sigma(\epsilon_k) = \Sigma J_A(\epsilon_k)$ includes the spectra of elements presented in Fig.1¹⁹⁾.

One can see that calculated spectra for all elements equally well fits the experiment at $\epsilon_k \lesssim 10^{14} \text{ eV}$ for both considered ISM phases. The maximum energy in the all particle spectrum $\epsilon_{max} \approx 10^{14} \text{ eV}$ and $\epsilon_{max} \approx 4 \times 10^{14} \text{ eV}$ for warm and hot ISM respectively only slightly exceeds the proton maximum energy.

One can expect that the observed CR spectrum which has the only peculiarity, so-called knee at $\epsilon \approx 3 \times 10^{15} \text{ eV}$, at energies $\epsilon \gtrsim 10^{15} \text{ eV}$ is produced by some reacceleration process. In this case one need to form in SNRs CR spectrum up to $\epsilon_{max} \approx 3 \times 10^{15} \text{ eV}$ which is essentially higher then calculated one. To demonstrate how CR spectrum could look like at $\epsilon > \epsilon_{max}$ CR spectra calculated at $B_0 = 12 \mu\text{G}$ and extended towards higher energies according to the law $\epsilon^{-3.1}$ is presented in Fig.1. This rather formal procedure gives the prediction of CR composition at energies $\epsilon_k \gtrsim 10^{15} \text{ eV}$, which is expected to be sensitive to the value ϵ_{max} .

According to the theoretical prediction the knee energy in different element spectra is proportional to their charge number Z and the mean CR atomic number increases roughly by a factor of two within the energy range $10^{15} \div 10^{16} \text{ eV}$. Preliminary experimental results of the ground-based installation KASCADE confirm these

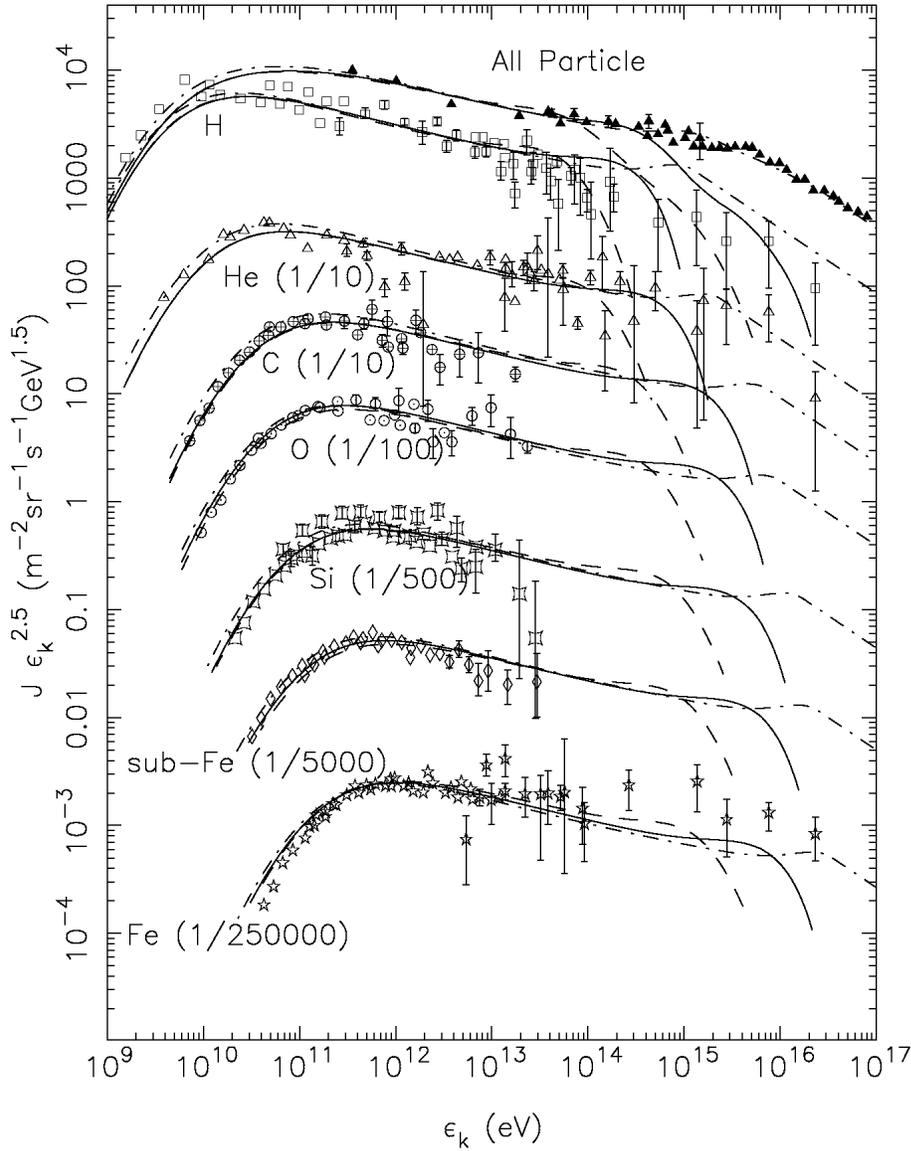


Figure 1: CR intensity near the Earth as function of the kinetic energy. Experimental points are taken from ²²). Solid (dashed) lines correspond to calculation for hot (warm) ISM with injection rate $\eta = 10^{-4}$. Dot-dashed lines correspond to hot ISM with magnetic field $B_0 = 12 \mu\text{G}$ and $\eta = 5 \times 10^{-4}$.

predictions²⁰⁾.

Due to extremely hard CR spectrum inside SNRs, predicted by the kinetic model, an essential contribution from the single nearby SNR ($d \lesssim 1$ kpc) should be observed as a kind of bump in galactic CR spectrum at energies $\epsilon = 10^{14} \div 10^{16}$ eV, if CRs leaking from the parent SNR expands into the galactic volume more or less spherically symmetric¹⁹⁾. It is not excluded that peculiarities in the galactic CR spectrum discussed by Erlykin and Wolfendale²¹⁾ can be attributed to this kind of effect.

4 Gamma-ray emission produced by nuclear component of CRs inside SNRs

Direct information about the dominant nucleonic CR component in SNRs can be obtained from γ -ray observations. If this nuclear component is strongly enhanced inside SNRs, then through nuclear collisions leading to pion production and subsequent decay γ -rays will be produced at detectable level⁹⁾.

The spectra of π^0 -decay γ -rays, produced by shock accelerated CRs in SNRs that expand into a the uniform ISM, that is typical situation for the case of SN Ia, were studied in detail in a kinetic approach¹¹⁾. We concentrate here on TeV γ -rays, measurable by the imaging Cherenkov technique.

The typical time-dependence of the expected integral flux $F_\gamma(\epsilon_\gamma)$ of γ -rays with energies ϵ_γ greater than 1 TeV from SNR Ia situated at 1 kpc distance in so-called warm ISM is shown in Fig.2. The predicted peak value of the expected γ -ray flux is

$$F_\gamma^{max}(1 \text{ TeV}) \approx 10^{-10} \left(\frac{N_H}{0.5 \text{ cm}^{-3}} \right) \frac{\text{photons}}{\text{cm}^2 \text{ s}}.$$

According to Fig.2 TeV γ -ray flux $F_\gamma > 10^{-11} \text{ cm}^{-2}\text{s}^{-1}$ is expected during about $t_m = 10^5$ yr of SNR evolution. It can be detected by the instrument with threshold $F_l \sim 10^{-12} \text{ cm}^{-2}\text{s}^{-1}$ up to a distance $d \approx 3$ kpc. Therefore taking into account that the expected number of SNRs of this kind $N_{sn} = \nu_{sn} t_m$ is about $N_{sn} \sim 300$ (the Galactic SN Ia rate is $\nu_{sn} \approx 1/300 \text{ yr}^{-1}$) one can conclude that about 10 SNRs of this type should be observable at any given time, by the imaging Cherenkov telescopes.

Application of the kinetic model to the case of SN 1006 gives some evidence that CR nuclear component provides the essential contribution in the observed TeV γ -ray flux²⁴⁾. Calculated synchrotron flux and integral γ -ray flux due to inverse-Compton (IC) scattering of CR electrons on microwave background radiation and due to collisions of the nuclear CR component with the gas nuclei presented in Fig.3 correspond to $E_{sn} = 10^{51}$ erg, $N_H = 0.1 \text{ cm}^{-3}$, $B_0 = 9 \mu\text{G}$, $\eta = 5 \times 10^{-4}$ and CR

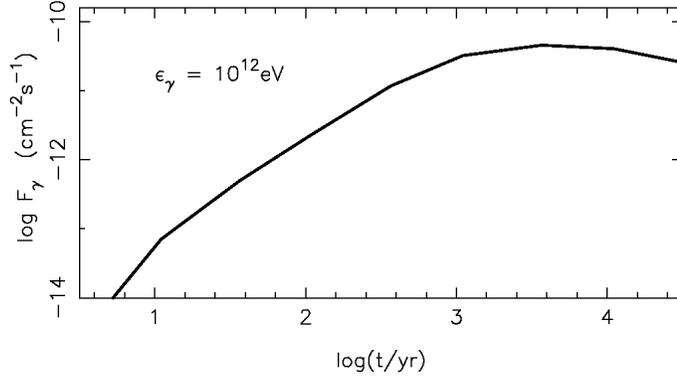


Figure 2: *Integral 1 TeV γ -ray flux expected at 1 kpc distance from SNR Ia situated in a warm ISM as a function of time since the SN explosion.*

electrons to protons ratio at relativistic energies 2×10^{-3} . One can see that the theory satisfactory reproduces observed synchrotron emission in radio ²⁵⁾ and X-ray ²⁶⁾ ranges and also fits existing γ -ray data ²⁷⁾. It is important to note, that contrary to simplified estimations ²⁸⁾ π^0 -decay γ -rays generated by nuclear CR component (p-p) and IC γ -rays generated by CR electrons give roughly equal contribution in the observed TeV γ -ray flux of SN 1006 (see also ²⁹⁾, where such possibility was discussed).

SNe of type Ib and II, which are more numerous in our Galaxy, explode into an inhomogeneous circumstellar medium, formed by the intensive wind of their massive progenitor stars ²³⁾.

The strong wind from the massive progenitor star interacts with an ambient interstellar medium (ISM) of uniform density $\rho_0 = 1.4mN_H$, resulting to first approximation in an expanding spherical configuration, which is called a bubble ²³⁾. Throughout its evolution, the system consists of four distinct zones. Starting from the center they are: (a) the hypersonic stellar wind (b) a region of shocked stellar wind (c) a shell of shocked interstellar gas, and (d) the ambient ISM.

Numerical results ¹²⁾ show that when a SN explodes into a circumstellar medium strongly modified by a wind from a massive progenitor star, then CRs are accelerated in the SNR almost as effectively as in the case of a uniform ISM: about 20 ÷ 40% of the SN explosion energy is transformed into CRs during the active SNR evolution.

During SN shock propagation in the supersonic wind region ($t \lesssim 10^3$ yr) very

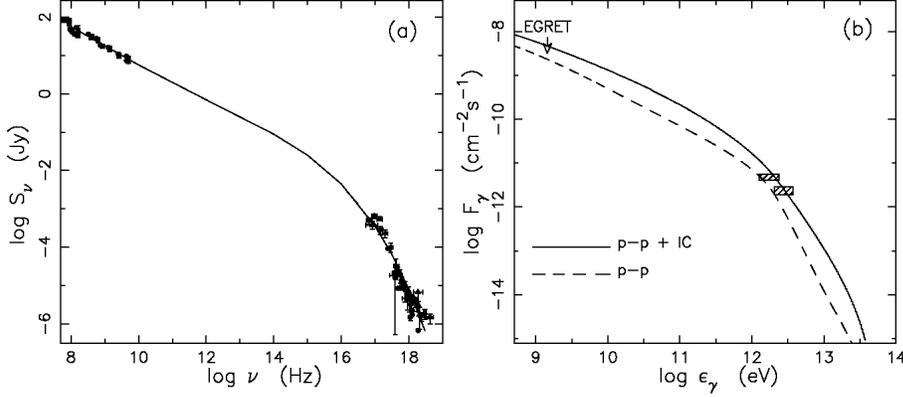


Figure 3: *Synchrotron flux as a function of frequency ν (a) and integral γ -ray flux as a function of energy ϵ_γ (b) from SN 1006.*

soon the acceleration process reaches a quasistationary level which is characterized by a high efficiency and a correspondingly large shock modification. Despite the fact that the shock modification is much stronger than predicted by a two-fluid hydrodynamical model ⁷⁾, the shock never becomes completely smoothed by CR backreaction: a relatively strong subshock with compression ratio always exists.

Due to the relatively small mass contained in the supersonic wind region CRs absorb there only a small fraction of the explosion energy (about 1% in the case of a SN type Ib, and 10% in the case of a SN type II) and the SNR is still very far from the Sedov phase after having swept up this region. Therefore we conclude, that the CRs produced in this region should not play a very significant role for the formation of the observed Galactic CR energy spectrum.

The peak value of the CR energy content in the SNR is reached when the SN shock sweeps up an amount of mass roughly equal to several times the ejected mass. This takes place during the SN shock propagation in the modified bubble.

The CR and γ -ray spectra are more variable during the SN shock evolution than in the case of a uniform ISM. At the same time the form of the resulting overall CR spectrum is rather insensitive to the parameters of the ISM as in the case of uniform ISM ¹²⁾. The maximum energy of the accelerated CRs reached during the SNR evolution is about 10^{14} eV for protons in all the cases considered, if the CR diffusion coefficient is as small as Bohm limit.

In the case of a SN Ib the expected TeV-energy γ -ray flux, normalized to a

distance of 1 kpc, remains lower than $10^{-12} \text{ cm}^{-2}\text{s}^{-1}$ during the entire SNR evolution if the ISM number density is less than 1 cm^{-3} except for an initial short period $t < 100 \text{ yr}$ when it is about $10^{-11} \text{ cm}^{-2}\text{s}^{-1}$. Only for a relatively dense ISM with $N_H = 30 \text{ cm}^{-3}$ the expected γ -ray flux is about $10^{-10} \text{ cm}^{-2}\text{s}^{-1}$ at late phases $t > 10^4 \text{ yr}$. A similar situation exists at late phases of SNR evolution in the case of SN II. The expected γ -ray flux is considerably lower, at least by a factor of hundred, compared with the case of uniform ISM of the same density N_H .

In the case of a SN II during the first several hundred years t_m after the explosion, the expected TeV γ -ray flux at a distance $d = 1 \text{ kpc}$ exceeds the value $10^{-9} \text{ cm}^{-2}\text{s}^{-1}$ and can be detected up to the distance $d_m = 30 \text{ kpc}$ with present instruments like HEGRA, Whipple or CAT. This distance is of the order of the diameter of the Galactic disk. Therefore all Galactic SNRs of this type whose number is $N_{sn} = \nu_{sn} t_m$ should be visible. But in this case we can expect at best $N_{sn} \sim 10$ such γ -ray sources at any given time.

An interesting example of type II SN represents SN 1987A in the Large Magellanic Cloud, because there are a lot of reliable observational data (e.g. see review³⁰⁾) which provide a unique opportunity to apply the existing models of CR and γ -ray production inside SNRs. In Fig.4a calculated time-dependence of SN shock size and speed³¹⁾ are compared with the experimental data³²⁾. During the initial period 1500 days after the explosion the SN shock propagated through the wind of BSG star which was a progenitor of SN 1987A during the last 10^4 yr before the SN event³⁰⁾. Due to low density of BSG wind the shock speed was very high $V_s \approx 30000 \text{ km/s}$ and almost constant. The essential shock deceleration from day 1500 to day 3000 indicates that the shock enters much more denser region occupied by the wind of RSG star: according to calculation the number density $N \approx 400 \text{ cm}^{-3}$ is required to reproduce the observed SN shock deceleration. Very high RSG wind number density leads to extremely high γ -ray luminosity on the current stage, as it is demonstrated in Fig.4b, despite of the large distance $d = 50 \text{ kpc}$. One can see from Fig.4b that expected γ -ray flux at $\epsilon_\gamma \lesssim 1 \text{ TeV}$ can be detected either by GLAST or by HESS instrument in the nearest future.

It is expected that the oldest SNRs which still confine accelerated CRs essentially contribute to the background diffuse galactic γ -ray flux. According to the estimations³³⁾ old SNRs as unresolved sources increase the expected TeV-energy γ -ray flux from the galactic disk by almost an order of magnitude. Therefore the measurements of the predicted diffuse galactic γ -ray flux at TeV-energies would give indirect

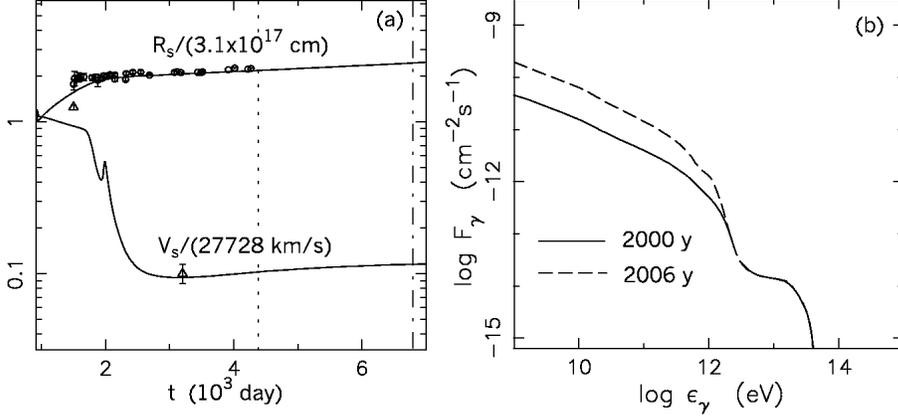


Figure 4: Shock size R_s and shock speed V_s (a) and integral γ -ray flux of SN 1987A expected at two different epochs (b): in 2000 (full line) and in 2006 (dashed line).

confirmation that SNRs are indeed the main sources of galactic CRs.

5 Summary

Detailed consideration performed within a frame of nonlinear kinetic model demonstrates, that the diffusive acceleration of CRs in SNRs is able to generate the observed CR spectrum up to an energy $10^{14} \div 10^{15}$ eV.

Due to relatively hard CR spectrum inside SNRs measurable contribution of nearby SNRs in the galactic CR spectrum seems to be quite probable.

According to the theoretical prediction about 20 SNRs should be visible in TeV γ -rays whereas only two were detected up to now, SN 1006²⁷⁾ and Cas A³⁴⁾. Negative correlation between SN of type Ia and ISM density could be a possible explanation of this deficit in detected TeV γ -ray sources. For core collapse SN of types II or Ib with quite massive progenitor one can in part explain this fact by the extremely low γ -ray intensity expected from such SNRs during the period of SN shock propagation through the low-density hot bubble. An alternative possibility relates to the confinement time which high energy CRs spend inside SNRs before their release into the interstellar medium: if it is essentially lower than 10^5 yr, that proportionally decrease the expected γ -ray sources. It is not also excluded that the spherical model overestimates CR acceleration efficiency in actual SNRs where particle injection/acceleration

can be suppressed at some essential part of the shock surface due to magnetic field structure. It can essentially reduce the expected number of SNRs visible in TeV γ -rays.

If SNRs produce CRs as effectively as predicted by the kinetic model then high energy diffuse galactic γ -ray flux is dominated by contribution of CRs situated inside old unresolved SNRs. The measurements of the diffuse flux at TeV-energies would give an indirect test whether SNRs are indeed the main source of CRs.

6 Acknowledgments

This work has been supported in part by the Russian Foundation of Basic Research grant 00-0217728 and by the Federal Scientific Program 'Astronomy' (grant 1.2.3.6). The author thanks the organizing committee for the invitation to present this paper and for the sponsorship.

References

1. L.O'C. Drury, Rep. Prog. Phys. **46**, 973 (1983).
2. R.D. Blandford and D. Eichler, Phys. Rept. **154**, 1 (1987).
3. E.G. Berezhko and G.F. Krymsky, Soviet Phys. Uspekhi. **12**, 155 (1988).
4. L.O'C. Drury, W.J. Markiewicz and H.J. Völk, A&A **225**, 179 (1989).
5. W.J. Markiewicz L.O'C. Drury and H.J. Völk, A&A **236**, 487 (1990).
6. E.A. Dorfi, A&A **234**, 419 (1990).
7. T.W. Jones and H. Kang, ApJ **396**, 575 (1992).
8. E.A. Dorfi, A&A **251**, 597 (1991).
9. L.O'C. Drury, F.A. Aharonian and H.J. Völk, A&A **287**, 959 (1994).
10. E.G. Berezhko, V.K. Yelshin and L.T. Ksenofontov, Astropart. Phys. **2**, 215 (1994); JETPh **82**, 1 (1996).
11. E.G. Berezhko and H.J. Völk, Astropart. Phys **7**, 183 (1997); **14**, 201 (2000).
12. E.G. Berezhko and H.J. Völk, A&A **357**, 283 (2000).

13. E.G. Berezhko, *Astropart. Phys.* **5**, 367 (1996).
14. L.O'C. Drury, H.J. Völk and E.G. Berezhko, *A&A* **299**, 222 (1995).
15. K.J. Trattner *et al.*, *JGR* **99**, 389 (1994).
16. K.B. Quest, *JGR* **93**, 9649 (1988).
17. M.A. Malkov and H.J. Völk, *A&A* **300**, 605 (1995).
18. E.G. Berezhko and D.C. Ellison, *ApJ* **526**, 385 (1999).
19. E.G. Berezhko and L.T. Ksenofontov, *JETPh* **116**, 737 (1999).
20. H. Ulrich *et al.*, *Proc.27th ICRC, Hamburg, 2001, Vol.5*, p.88.
21. A.D. Erlykin and A.W. Wolfendale, *Astropart. Phys.* **7**, 1 (1997).
22. T. Shibata, *Proc. 24th ICRC, Rome 1995, Invited, Rapporteur & Highlight Papers*, p.713.
23. R. Weaver *et al.*, *ApJ* **218**, 377 (1977).
24. E.G. Berezhko, L.T. Ksenofontov and S.I. Petukhov, *Proc. 26th ICRC, Salt Lake City 1999, Vol.4*, p.431.
25. S.P.Reynolds, *ApJ* **459**, L13 (1996).
26. A.J.S. Hamilton, C.L. Sarazin and A.E. Szymkowiak, *ApJ* **300**, 698 (1986).
27. T. Tanimori *et al.*, *ApJ* **497**, L25 (1998).
28. A. Mastichiadis and O.C. De Jager, *A&A* **311**, L5 (1996).
29. F.A. Aharonian and A.M. Atoyan, *A&A* **351**, 330 (1999).
30. R. Mc Cray, *Ann. Rev. Astron. Astrophys.* **31**, 175 (1993).
31. E.G. Berezhko and L.T. Ksenofontov, *Astron. Lett.*, **26**, 639 (2000).
32. B.M. Gaensler *et al.*, *ApJ* **479**, 845 (1997).
33. E.G. Berezhko and H.J. Völk, *ApJ.*, **540**, 923 (2000).
34. F.A. Aharonian *et al.*, *A&A*, **370**, 112 (2001).

**THE INTERPRETATION OF RADIOACTIVE ISOTOPES IN
THE COSMIC RADIATION AND THE LINK BETWEEN THE
DIFFUSION HALO MODEL AND THE LEAKY BOX MODEL**

Manfred Simon

University of Siegen, Physics Department, 57068 Siegen, Germany

ABSTRACT

This article deals with the propagation of cosmic ray particles in the galaxy. It describes the physical scenario of the Leaky Box Model (LBM) and the Diffusion Halo Model (DHM), presents the appropriate equations and their solutions and compares them with the data. The aspect is stressed that the DHM is the proper physical model in which the propagation should be described while the LB calculation only follows as a mathematical approximation from the DHM. The secondary radioactive particles are used to separate the diffusion coefficient $D(E)$ and the halo size H , which are the physical parameters in the DHM.

1 Introduction

The elemental and isotopic composition of galactic cosmic rays (CR) contains a record of the nuclear history of a sample of matter from other regions of

the galaxy, including its synthesis in stars, its process of acceleration and the subsequent transport through the interstellar medium.

It is currently believed that the cosmic ray particles at least up to about 10^{15} eV are of galactic origin, accelerated by shock waves of expanding supernova remnants (SNR). Direct evidence for the acceleration of particles in the vicinity of the supernova remnants (SNR) ¹⁾, comes from TeV- γ observations ¹⁸⁾. It is, however, not well understood what type of particles (electrons and/or hadrons) and which fraction of matter is actually selected from the ambient plasma for acceleration and finally injected as energetic particles into the interstellar space, where they spend about 10^7 years before they finally escape into the intergalactic space. This long confinement time suggests a coupling to the galactic magnetic fields by Lorentz forces since the residence time along a line of sight through our galaxy would only be 10^3 years at most. In addition, the high degree of isotropy which we observe in the cosmic radiation suggests an effective diffusion since cosmic rays streaming freely out of the galaxy or following strictly the galactic magnetic field pattern would be highly anisotropic with most flux coming from the direction of the central region of the galaxy. The spectrum of magnetic field irregularities may serve as scatter center in this diffusion process.

The relatively long confinement time allows the CR particles to interact in various ways with the constituents of the interstellar environment such as gas, magnetic fields or photons. These high energy interactions lead to the production of new particles, new photons of different wave lengths (e.g. radio, microwave, IR, optical, UV, X- and γ -rays) and also to spallation products (often called "secondary" nuclei) which result when the more abundant "primary" heavier nuclei break up while they interact with the interstellar gas, thus from a variety of different observations (radio, high energy photons, secondary/primary ratios (s/p ratio) and cosmic ray spectra) we have accumulating evidence that the existence of cosmic ray particles is a general feature everywhere in our galaxy.

We also know from the distribution of synchrotron emission which occurs when high energy electrons spiral around magnetic field lines and which we observe from edge-on spiral galaxies such as NGC 891 that the high energy particles do not strictly restrain to the thin galactic disk but propagate out into the halo. Thus, the volume in which cosmic rays can be found is larger than that given by the thin galactic disk where most of the stars and energetic processes take

place, and where probably also the cosmic ray sources are located.

In the framework of these observational constraints models of CR propagation have been developed and there are actually two which are mostly referred to in the literature. It is the Leaky Box Model (LBM) and the Diffusion Halo Model (DHM). Both models are able to explain the CR data surprisingly well although the physical framework of both models differs.

The LBM describes an equilibrium model, in which the cosmic ray sources, and the primary and secondary cosmic ray particles are homogeneously distributed in a confinement volume (box, galaxy) and constant in time with no gradient of CR density into any direction. Thus, the transport of CR is not controlled by diffusion but by a hypothetical leakage process at the imaginary boundaries. The mechanism which particles use to escape from this confinement volume is not addressed and only parametrized by a leakage or escape time τ_{escape} . This scenario leads to an equilibrium equation which balances the production and losses of particles.

Despite the fact that the LBM is very popular and often used in the literature the problems are obvious: What is the physics of this box? Which mechanism keeps the particles homogeneously distributed within this box? What is the mechanism of escape and last but not least we have strong evidence that the CR sources are located within the thin galactic disk and not equally distributed in a box larger than the disk.

The DHM, on the other hand, deals with a more realistic physical scenario, since it takes more observational physical constraints into consideration ¹¹⁾. In this model the sources of CR are indeed distributed within the thin galactic disk and the escape from the disk into the halo and finally into the intergalactic space is determined by diffusion. In this DHM one thus expects a gradient of CR density away from the galactic disk implying a constant streaming of CR particles away from the galactic disk into the halo. This streaming is determined by a diffusion coefficient $D(E)$ and the halo size H .

The secondary, radioactive CR nuclei with decay times comparable with the confinement time play an important role in determining these physical parameters and also provide information on the mean gas density through which the particles propagate.

In this article I will show how this works and will present and discuss the relevant equations of both models, i.e. the LBM and the DHM. I will present

their solutions and will compare the predictions with the data from CR observations. As a conclusion out of this discussion I will stress the position that the DHM provides the realistic physical scenario in which CR propagation should be treated. I will illustrate that the LB calculation follows as a mathematical approximation from the DHM. For that reason the LB calculations provide the same mathematical results. This works for stable CR nuclei, for radioactive particles, however, this approximation fails.

The reader is also referred to other books and reviews (3, 8, 19, 33).

2 Leaky Box Model

In the LBM the particles propagate freely in the containment volume and the productions and losses of particles are balanced in time, thus the mathematical description of the LBM is given by a continuity equation. By ignoring energy changing processes and radioactive particles the equation becomes the following form:

$$N_i(E) \left\{ \frac{1}{\tau_{\text{esc}}(E)} + \frac{1}{\tau_{\text{int}}(E)} \right\} = {}_iQ_{\text{prim}}(E) + \sum_{k>i} \frac{N_k(E)}{\tau_{\text{int}}^{k \rightarrow i}(E)} \quad (1)$$

where $N_i(E)$ [$\text{cm}^{-3} \text{ GeV}^{-1}$] and $N_k(E)$ [$\text{cm}^{-3} \text{ GeV}^{-1}$] stands for the number densities of different types of nuclei of kinetic energy E . The left side of this equation (1) accounts for the losses of i -type nuclei and the right side for the sources. The secondary sources originate from spallation of k -type nuclei heavier than the i -type nuclei. The quantity $\tau_{\text{int}}^{k \rightarrow i}(E)$ means the mean time which a k -type nuclei needs to produce an i -type secondary in the interstellar gas. This quantity depends on the production cross-section and the interstellar gas in terms of density and composition. As one can see a huge number of nuclear cross-sections are involved and not all are so well-known. For more details on these cross-sections I refer to the literature (15, 16, 31). In the brackets on the left side one finds expressions for the losses and $\tau_{\text{int}}(E)$ stands for the mean lifetime of i -type particles against interactions in the interstellar gas. Here again nuclear cross-sections are involved. In the LBM the quantity $\tau_{\text{esc}}(E)$ is used as a free parameter and stands for the mean escape time from the confinement volume, often called the age of cosmic rays. On a statistical basis, however, it is an exponential distribution which governs the escape of an individual particle and $\tau_{\text{esc}}(E)$ is the mean of it. The exponential distribution results since the probability for a particle to escape from the box in time dt is

given by $dt/\tau_{\text{esc}}(E)$.

One often finds the characteristic times $\tau_{\text{esc}}(E)$, $\tau_{\text{int}}(E)$, $\tau_{\text{int}}^{k \rightarrow i}(E)$ replaced by lambdas which characterize the matter traversed in $[\text{g}/\text{cm}^2]$. These relations are given by:

$$\lambda_{\text{esc}}(E) = \langle m \rangle \cdot \bar{n}(\text{cm}^{-3}) \cdot c \cdot \beta \cdot \tau_{\text{esc}}(E) \quad (2)$$

$$\lambda_{\text{int}}(E) = \langle m \rangle \cdot \bar{n}(\text{cm}^{-3}) \cdot c \cdot \beta \cdot \tau_{\text{int}}(E) \quad (3)$$

$$\lambda_{\text{int}}^{k \rightarrow i}(E) = \langle m \rangle \cdot \bar{n}(\text{cm}^{-3}) \cdot c \cdot \beta \cdot \tau_{\text{int}}^{k \rightarrow i}(E) \quad (4)$$

where $\bar{n}(\text{cm}^{-3})$ refers to the mean interstellar gas density through which the particles penetrate, $\langle m \rangle$ means the mean mass of the gas and $c \cdot \beta$ is the velocity of the particle. If one further relates the lambdas $\lambda_{\text{int}}(E)$ and $\lambda_{\text{int}}^{k \rightarrow i}(E)$ to their cross-section $\sigma_{\text{int}}(E)$ and $\sigma_{\text{int}}^{k \rightarrow i}(E)$ via

$$\lambda_{\text{int}}(E) = \frac{\langle m \rangle}{\sigma_{\text{int}}(E)} \quad (5)$$

$$\lambda_{\text{int}}^{k \rightarrow i}(E) = \frac{\langle m \rangle}{\sigma_{\text{int}}^{k \rightarrow i}(E)} \quad (6)$$

one can write eq. (1) in the form:

$$\frac{N_i(E)}{\tau_{\text{esc}}(E)} + \bar{n} \cdot v \cdot \sigma_{\text{int}}(E) \cdot N_i(E) = q_i + \sum_{k>i} \bar{n} \cdot v \cdot \sigma_{\text{int}}^{k \rightarrow i}(E) \cdot N_k \quad (7)$$

For practical purpose one has to rearrange eq. (1) and (7) since one measures fluxes $I[\text{cm}^{-2}\text{s}^{-1}\text{GeV}^{-1}\text{ster}^{-1}]$ and not densities $N[\text{cm}^{-3}\text{GeV}^{-1}]$. This relation can be found since we assume isotropy in the cosmic radiation:

$$N(E) = \frac{4\pi}{v} I(E) \quad (8)$$

where v is the velocity of the particle.

The above equilibrium eq. (1) and (7) can be solved by different mathematical techniques and I refer to the literature (12, 14). I like to note that care has to be taken when energy changing processes are involved (9, 10, 14, 29). Solutions of eq. (1) or (7) provide predicted abundances for the different cosmic ray species and fits to the observed cosmic ray spectra and to a secondary to primary ratio, such as B/C, allow to determine the matter traversed by the particles, $\lambda_{\text{esc}}(E)$, and the relative abundance of the primary particles at their

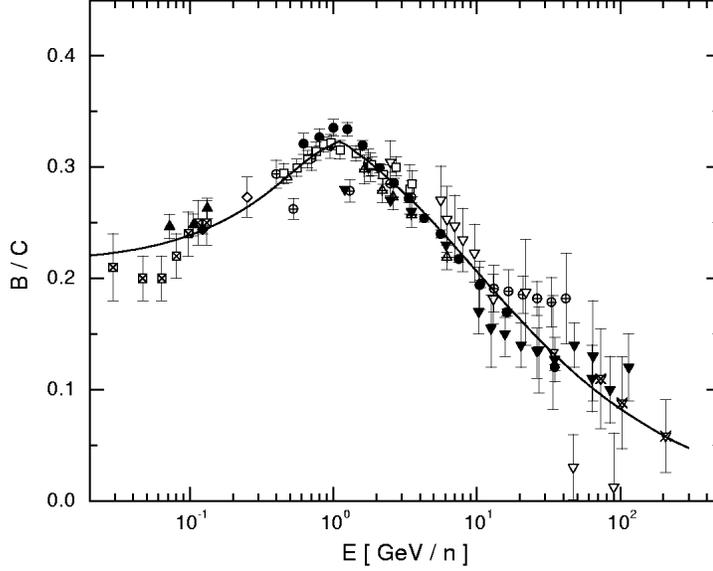


Figure 1: Collection of measured B/C ratios at different energies ^{10, 17}). The curve is a fit to the data. This fit determines the λ_{esc} -dependence in the LBM, see eq. (9), and in the DHM it settles the ratio of D/H , see. eq. (14).

sources ¹⁴).

In Fig. 1 I show, as an example, the fit to the measured B/C -ratio and eq. (9) describes the corresponding $\lambda_{\text{esc}}(R)$ dependence as a function of rigidity ²¹)

$$\lambda_{\text{esc}}(R) = \begin{cases} 12.8 [g/cm^2] \left(\frac{R}{4.7 \text{ GV}}\right)^{0.8} & \text{for } R < 4.7 \text{ GV} \\ 12.8 [g/cm^2] \left(\frac{R}{4.7 \text{ GV}}\right)^{0.57} & \text{for } R > 4.7 \text{ GV} \end{cases} \quad (9)$$

Similar calculations have been done by many authors and Fig. 2 shows some of the recently published curves on the rigidity dependence of $\lambda_{\text{esc}}(R)$, which are used by the various authors to fit the data on cosmic rays. Although there are quite remarkable differences between these curves, particularly at low energies, which may be partly due to cross-section uncertainties, they agree in their

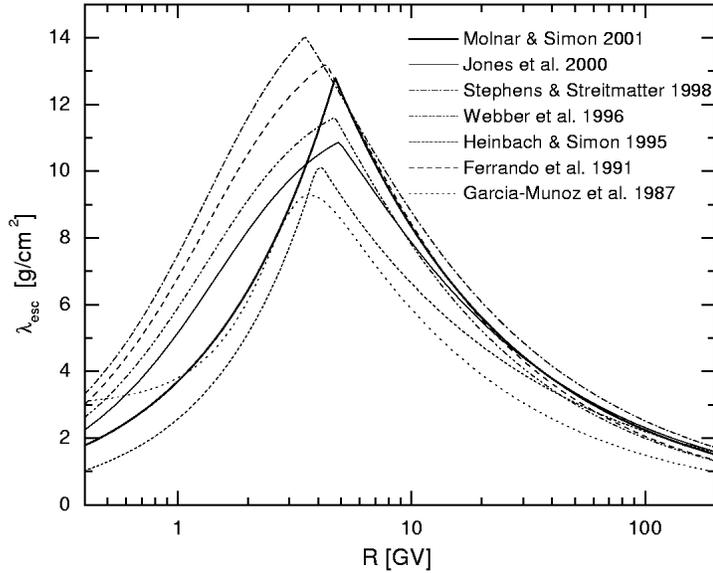


Figure 2: Rigidity dependences of the mean escape length $\lambda_{\text{esc}}(E)$ as published by different authors. This quantity is a free parameter in the LB calculation and results by fitting measured secondary/primary-ratios, such as B/C , see Fig. 1.

general shape.

They all peak around some GeV/nucleon and fall off to higher and lower energies. These curves tell us that the CR particles should traverse some (g/cm^2) interstellar matter – around 1 GeV/n it is roughly $10(\text{g}/\text{cm}^2)$ – before they escape from the confinement volume. A plausible physical reason for this characteristic shape does not exist and a physical explanation is still open. I will mention that this λ_{esc} -dependence which fits the CR data is sensitive to the physical condition under which the propagation occurs. For instance, if one allows the particles to gain energy during propagation, then one can show that the λ_{esc} -dependence stands as a single power law in rigidity over the whole energy regime and the peak around 1 GeV/n disappears. For more details on this reacceleration aspect I refer to the literature (14, 22, 25, 27).

Various calculations on CR data have been done in the framework of this LBM and one obtained always good agreement with the CR data. This applies to the cosmic ray spectra (6, 14) to the antiprotons (9, 28) and to the positrons (22, 24). This ability gave credit to this LBM although its physical picture is very questionable. In the next chapter I will introduce the DHM which I consider more as a realistic physical model within which one should discuss the propagation of cosmic rays. I will also show that the good agreement between the cosmic ray data and the LB calculations does not come as a surprise since the LB equations represent a mathematical approximation of the DHM.

3 The Diffusion Halo Model

Fig. 3 sketches the physical picture of the DHM. The shaded area illustrates the thin galactic disk of height h and the quantity H stands for the height of the halo. It is assumed that the cosmic ray sources are placed in the thin galactic disk, where most of the interstellar gas is located but the cosmic rays themselves diffuse out and may spend considerable portions of their lifetime in the halo. A three dimensional diffusion equation would be the proper approach to the problem. It could cover physical details in our galaxy such as the spatial gas distribution of atomic and molecular hydrogen, could link the cosmic ray sources to the distribution of supernova remnants (SNR), and could also add aspects such as galactic winds and convection (30). This readily illustrates that the DHM accounts for much more physical detail than the LBM.

For our discussion here it is sufficient to make the picture somehow simpler. I will allow that the cosmic ray sources and the interstellar gas are homogeneously distributed throughout the thin galactic disk and will ignore convection and energy changing processes. This provides symmetry and if one further ignores energy changing processes one can describe this scenario with a one-dimensional diffusion equation:

$$\begin{aligned} \frac{\partial N_i(z, t)}{\partial t} = \frac{\partial}{\partial z} \left\{ D(z) \frac{\partial}{\partial z} N_i(z, t) \right\} - N_i(z, t) \left\{ \frac{1}{\tau_{\text{int}}(E)} \right\} + \\ + {}_i Q_{\text{prim}}(z) + \sum_{k>i} \frac{N_k(z, t)}{\tau_{\text{int}}^{k \rightarrow i}}. \end{aligned} \quad (10)$$

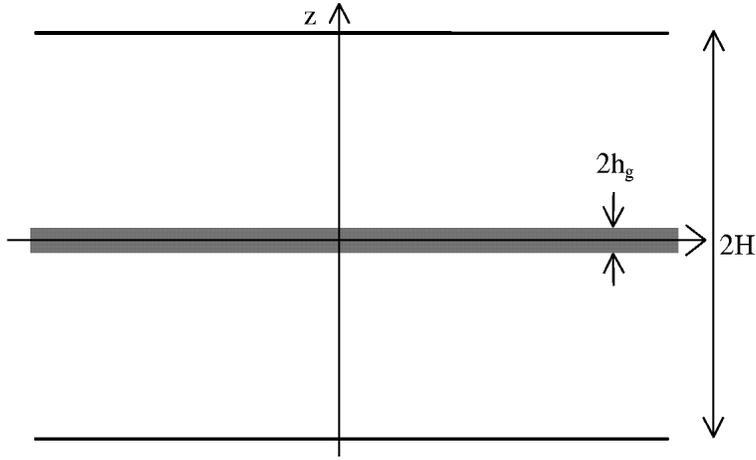


Figure 3: *Schematic view of the physical concept of the DHM. The shaded area symbolizes the thin galactic disk and H stands for the halo size. The sources of cosmic rays are in the disk and the escape into the halo and finally into the intergalactic space is controlled by diffusion.*

Various quantities and terms can be found. $N_i(z, t)$ and $N_k(z, t)$ describe the density of i -type and k -type particles at position z at time t with k heavier than i . The first term of the right side describes the diffusion and $D(z)$ means the diffusion coefficient at position z . For simplicity I allow D to be independent of position. The second bracket on the right side of eq. (10) accounts for the losses of i -type particles similar to those quantities described in the last chapter. The first term in the bracket stands for the losses against interactions. In the last two terms one finds the sources for the i -type particles. One can have primary sources ${}_iQ_{\text{prim}}(z)$ as well as secondary sources by spallation of k -type nuclei expressed by the last term on the right side of eq. (10). Various interesting results can be obtained by solving eq. (10) under different sets of parameters and boundary conditions.

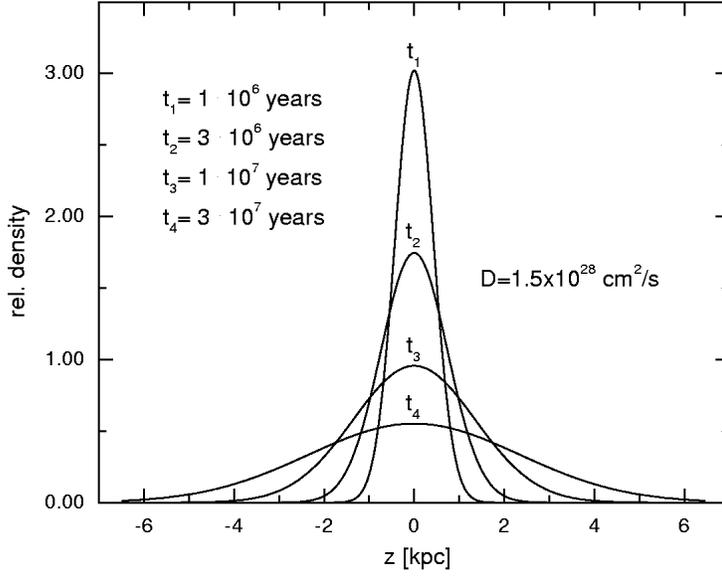


Figure 4: *Particles which are injected at the center line of the galactic disk ($z = 0$) diffuse out into the halo according to the diffusion coefficient D . These curves represent the probability of finding the particle at position z after a certain elapsed time.*

I will begin with Fig. 4. These curves solely illustrate the diffusion of particles out of the galactic disk into the halo as a function of time after being injected at $t = 0$ and $z = 0$ in a single shot. Losses and secondary productions as expressed on the right side of eq. (10) were ignored. The diffusion coefficient was chosen to $D = 1.5 \cdot 10^{28} \text{ cm}^2/\text{s}$.

These curves are represented by the following analytic formula:

$$N(z, t) = \frac{1}{2(\pi \cdot D \cdot t)^{1/2}} \cdot \exp \left\{ -\frac{z^2}{4Dt} \right\}. \quad (11)$$

They allow to derive the mean square of the displacement which the particles encounter on a statistical basis by diffusing away from the center line of the

galactic plane into the halo as a function of time. This is given by

$$\langle z^2 \rangle = \int_{-\infty}^{\infty} z^2 N(z, t) dz = 2Dt . \quad (12)$$

If one interprets $\sqrt{\langle z^2 \rangle}$ with the halo size H one obtains the mean time t_{esc} which a particle needs to escape freely into the intergalactic space:

$$t_{\text{esc}} = \frac{H^2}{2D} . \quad (13)$$

In order to calculate this confinement time in the framework of the DHM one needs information on the halo size H and the diffusion coefficient. This information, however, can be obtained directly from the data by using secondary/primary ratios, such as B/C , and secondary radioactive isotopes. This works as follows. Fig. 5 illustrates the equilibrium state for two stable particles, for the primary carbon and the secondary boron. Both curves were obtained by solving eq. (10) numerically in an iterative procedure until equilibrium is reached. We allowed a mean gas density n_d of 1 H-atom/cm³ in the thin galactic disk and in the halo we set it to zero ²¹). In Fig. 5 we also give the values for D and H which we used in this calculation. As can be seen in Fig. 5, the B/C ratio is position dependent. It is smaller in the halo than in the disk. This is very different to the LB situation where everything in the box is believed to be the same everywhere.

For comparison with the CR data one has to take the calculated B/C ratio at position $Z = 0$, where we actually measure the cosmic rays. One learns from these calculations, that the ratio of D/H actually determines the B/C -ratio. If one increases the D/H ratio the B/C ratio will also increase and vice versa. This results from a combination of the production of boron within the gaseous disk and its streaming away into the halo.

Thus a fit to the B/C -data can be obtained in the DHM by varying the D/H ratio and it can be shown that the following dependence ²¹):

$$\frac{D(R)}{H} = \begin{cases} 1.7 \cdot 10^6 [cm/s] \beta \left(\frac{R}{4.7 \text{ GV}}\right)^{-0.8} & \text{for } R < 4.7 \text{ GV} \\ 1.7 \cdot 10^6 [cm/s] \beta \left(\frac{R}{4.7 \text{ GV}}\right)^{0.57} & \text{for } R > 4.7 \text{ GV} \end{cases} \quad (14)$$

reproduces the same curve which is shown in Fig. 1. These identical results from the LB calculations and the DHM do not come as a surprise, it results from a mathematical relation between the DHM and the equilibrium equation

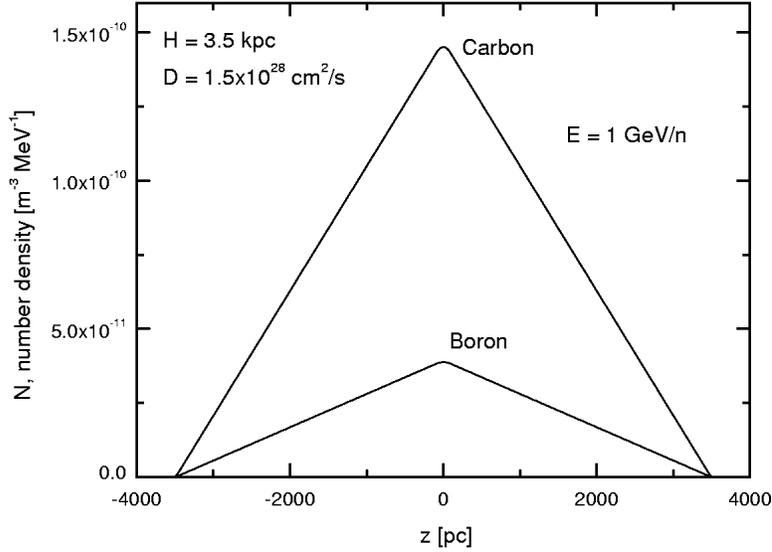


Figure 5: *Calculated equilibrium states in the framework of the DHM for the primary carbon and the secondary boron under the given parameters of H and D . From these calculations one finds that the ratio of D/H determines the B/C -ratio. An increase of D/H leads to an increase of B/C and vice versa. Thus, a fit to the measured B/C -ratio (measured at position $z = 0$) settles the D/H -ratio. See text and eq. (14).*

which describes the LBM. This will be illustrated in the following.

If equilibrium is reached stable CR nuclei are described in the DHM by the following equation ²⁾:

$$D \frac{\partial^2}{\partial z^2} N_i(z) - n(z) \cdot v \cdot \sigma_i \cdot N_i(z) - q_i(z) = 0 \quad (15)$$

All these quantities are explained above and $q_i(z)$ is the source term for primary and secondary stable nuclei. One can derive this eq. (15) from eq. (10) by using the relations given in eqs. (2–6). Restricting the gas $n(z)$ and the sources of CR to the disk, via

$$n(z) = n_d \theta(h - |z|) \quad (16)$$

$$q_i(z) = q_{i,d}\theta(h - |z|), \quad (17)$$

where θ stands for the Heavyside function, eq. (15) has an analytic solution 2):

$$N_i(z) = \frac{q_{i,d}}{n_d \cdot v \cdot \sigma_i} \{1 - \cosh[\lambda_g \cdot z][\cosh(\lambda_g \cdot h) + \lambda_g(H - h) \sinh(\lambda_g \cdot h)]^{-1}\}$$

with

$$\lambda_g = \sqrt{\frac{n_d \cdot v \cdot \sigma_i}{D}} \quad (18)$$

Under the conditions: $h \ll H$ and $\lambda_g \cdot h \ll 1$ (which are normally fulfilled by the given quantities) one obtains for $z = 0$ the following approximation:

$$N_i \approx \frac{q_{i,d}}{n_d \cdot v \cdot \sigma_i} \{1 - [1 + \lambda_g(H - h) \cdot \lambda_g \cdot h]^{-1}\} \quad (19)$$

Hence

$$N_i \approx q_{i,d} \cdot \frac{h \cdot H/D}{1 + n_d \cdot v \cdot \sigma_i \cdot h \cdot H/D} \quad (20)$$

If we now invent the following equation:

$$\frac{N_i}{\tau_{\text{esc}}} + \bar{n} \cdot v \cdot \sigma_i \cdot N_i = \bar{q}_i \quad (21)$$

with $\tau_{\text{esc}} = \lambda_{\text{esc}}/(\bar{n} \cdot \langle m \rangle \cdot v)$, $\bar{q}_i = (\bar{n}/n_d) \cdot q_{i,d}$ (these equations are explained above)

and compare its solution:

$$N_i = \bar{q}_i \cdot \frac{\tau_{\text{esc}}}{1 + \bar{n} \cdot v \cdot \sigma_i \cdot \tau_{\text{esc}}} = q_{i,d} \cdot \frac{\lambda_{\text{esc}}/(n_d \cdot \langle m \rangle \cdot v)}{1 + \sigma_i \cdot \lambda_{\text{esc}}/\langle m \rangle} \quad (22)$$

with eq. (20) one realizes that both solutions are equal for

$$\lambda_{\text{esc}} = \frac{n_d \cdot \langle m \rangle \cdot h \cdot H \cdot v}{D} \quad (23)$$

The eq. (21) which we invented, however, is the LB equation which we used in the previous chapter. Thus, the LB equation and an approximation of the DHM provide the same mathematical result for stable CR nuclei and the link is given by equation (23). In this interpretation the LBM has not to be taken as a realistic physical model of CR propagation. It is a mathematical approximation and works only for stable nuclei and fails for the case of secondary radioactive nuclei. In this view the LBM cannot and should not be used for an interpretation of the radioactive secondary isotopes such as

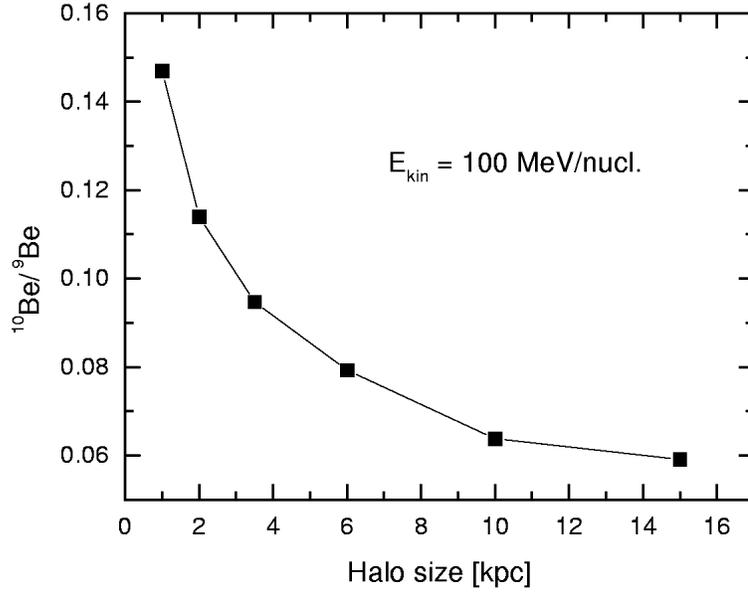


Figure 6: *In the DHM the fit to secondary/primary-ratios only settles the D/H-ratio. In order to disentangle D and H one has to make use of secondary, radioactive nuclei such as Be10. The curve illustrates how the calculated Be10/Be9-ratio changes as a function of the halo size H. This calculation was done under the condition of a constant D/H ratio, $D/H = 2.3 \cdot 10^6$ cm/sec, which ensures an unchanged production of the stable Be9 isotope.*

Be-10	$\tau_d = 2.3 \cdot 10^6$ years
Al-26	$\tau_d = 1.0 \cdot 10^6$ years
Cl-36	$\tau_d = 4.5 \cdot 10^5$ years
Mn-53	$\tau_d = 5.4 \cdot 10^5$ years

The interpretation of these isotopes have to be done in the framework of DHM, which I will do next.

4 Diffusion Halo Model and Secondary Radioactive Cosmic Ray Nuclei

As shown in the previous section, in the DHM the B/C-ratio or any other secondary/primary ratio can only settle the ratio of D/H, and not D and H individually, because stable secondaries are sensitive to both parameters. Radioactive secondaries, however, are insensitive to the halo size H, since most of them are already decayed before they reach this boarder by diffusion. Thus, measurements on radioactive secondaries like those mentioned above provide the mean to separate D and H in the DHM. This is illustrated in Fig. 6. We plotted the calculated Be10/Be9 ratio as a function of the halo size H at an energy of 100 MeV/n. This calculation was performed under the condition of a constant D/H ratio, $D/H = 2.3 \cdot 10^6$ (cm/sec) which ensures a constant flux of the stable secondary Be-9 isotope. The decrease of this curve with D or H is thus caused by a decrease of the Be-10 flux. By comparing measured Be10/Be9-ratio with the calculation in combination with the proper D/H ratio which fits the secondary/primary-ratios one is able to deduce both parameters, D and H, individually.

In Fig. 7 I show measured ratios of Be10/Be9 along with calculated curves which refer to different halo sizes H. But in all these curves the ratio of D/H was the same and given by eq. (14) which fits the B/C-ratio. As one can see the low energy data on the Be10/Be9-ratio around 100 MeV/n agree with a halo size of $H = 3.5$ kpc and a diffusion coefficient of $D = 2.49 \cdot 10^{28}$ (cm²/sec) which results in an escape time of $\tau_{\text{esc}} = 7.6 \cdot 10^7$ years, see eq. (13). It becomes also obvious from Fig. 7 that the high energy data around 1 GeV/n, which stam from the recent balloon borne ISOMAX experiment, do not necessarily agree with this halo size of $H = 3.5$ kpc. These data favour more a smaller halo size, indicating that the effective halo size may decrease with the energy of the particle. That would mean that the volume around our galaxy which is populated with particles extend further out for the low energy particles than for the high energy particles. I regard this conclusion at this time as an interesting possibility, but more data with smaller error bars are needed to confirm this interpretation.

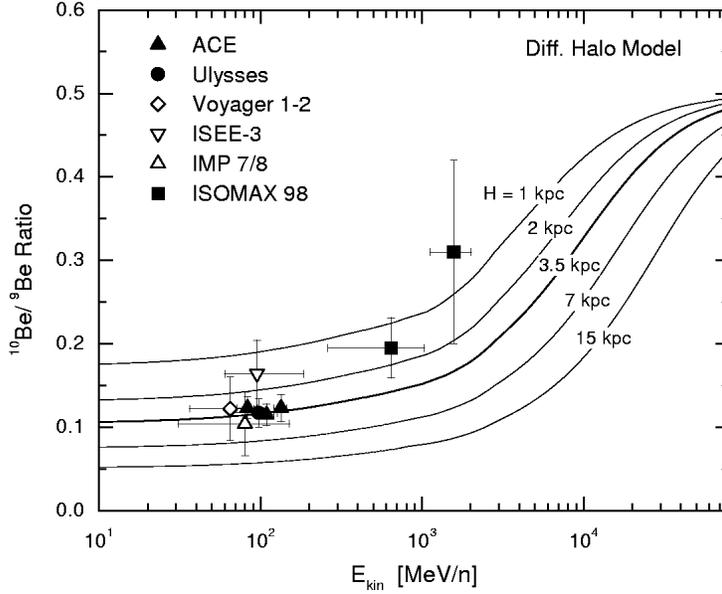


Figure 7: Measured Be_{10}/Be_9 -ratios at different energies. The curves represent calculations in the framework of the DHM. All these curves satisfy the D/H ratio as given by eq. (14), thus provide a good fit to the B/C -ratio. The low energy data agree with the halo size of $H = 3.5$ kpc while the high energy data favour more a smaller halo size, see text. Data: IMP7/8 ¹⁰⁾, ISEE-3 ³⁴⁾, Voyager 1-2 ²⁰⁾, Ulysses ^{4, 5)}, ISOMAX 98 ^{13, 23)} and ACE ³⁵⁾.

5 Acknowledgement

I wish to thank A. Molnar for many discussions on this topic and for his help in preparing this article.

References

1. W.J. Axford, Proc. 18th ICRC (Paris), **12**, 155 (1981).
2. V.S. Berezhinskii *et al.*, Astrophysics of Cosmic Rays (North Holland, Amsterdam, 1990).

3. C.J. Cesarsky, Rapporteur talk, 20th ICRC, **8**, 87 (1987).
4. J.J. Connell, 25th ICRC, **Vol. 3**, 385 (1997).
5. J.J. Connell, *Ap.J.*, **501**, L59, (1998).
6. J.J. Engelmann *et al.*, *A & A*, **96**, 233 (1990).
7. P. Ferrando *et al.*, **163**, 247 (1991).
8. T.K. Gaisser, *Cosmic rays and particle physics*, Cambridge University Press (1990).
9. T.K. Gaisser & R.K. Schaefer, *Ap.J.*, **174**, 394 (1992).
10. M. Garcia-Munoz *et al.*, *Ap.JS*, **64**, 269 (1987), 17th ICRC, **2**, 72 (1981).
11. V.L. Ginzburg, Ya.M. Khazan & V.S. Ptuskin, *Ap&SS*, **68**, 295 (1980).
12. V.L. Ginzburg & Syrovatskii, *The Origin of Cosmic Rays*, Pergamon Press (1964).
13. T. Hams *et al.*, *Proc. 27th ICRC*, OG1.1 (2001).
14. U. Heinbach & M. Simon, *Ap.J.*, **209**, 441 (1995).
15. W. Heinrich *et al.*, *Radiation Measurement*, **25**, 203 (1995).
16. J. Isbert *et al.*, 23th ICRC, **2**, 167 (1993).
17. F.C. Jones, A. Lukasiak, V. Ptuskin, and W. Webber, *Ap.J.*, **547**, 264 (2001).
18. D. Kranich *et al.*, 27th ICRC, 2630 (2001).
19. M.S. Longair, *High energy astrophysics*, Cambridge University Press (1981).
20. A. Lukasiak, P. Ferrando, F.B. McDonald & W.R. Webber, *Ap.J.*, **423**, 426 (1994).
21. A. Molnar & M. Simon, 27th ICRC, 1860 (2001).
22. J.V. Moskalenko and A.W. Strong, *Ap.J.*, **493**, 694 (1998).

23. G.A. deNolfo *et al.*, Proc. 27th ICRC, OG1.1 (2001).
24. R.J. Protheroe, Ap.J., **254**, 391 (1982).
25. E.S. Seo & V.S. Ptuskin, Ap.J., **431d**, 705 (1996).
26. M. Simon & U. Heinbach, Ap.J., **456**, 519 (1996).
27. M. Simon, W. Heinrich & K.D. Mathis, Ap.J., **300**, 32 (1986).
28. M. Simon, A. Molnar and S. Roesler, Ap.J., **499**, 250 (1998).
29. Stephens & Streitmatter, 20. Sept. Ap.J., 505 (1998).
30. A.W. Strong & J.V. Moskalenko, 509 (1998).
31. W.R. Webber, J.C. Kish & D.A. Schrier *et al.*, Phys.Rev. C **41**, 520 (1990).
32. W.R. Webber *et al.*, Ap.J., **457**, 435 (1996).
33. W.R. Webber, Space Science Reviews, **81**, 107 (1997).
34. M.E. Wiedenbeck & D.E. Greiner, Ap.J.Lett., **239**, L139 (1980).
35. N.E. Yanasak *et al.*, Proc. 26th ICRC, (1999).

**Propagation of galactic cosmic rays and antiprotons in a diffusion
model**

Fiorenza Donato
LAPTH, BP 110, Annecy-le-Vieux, 74941 France

ABSTRACT

Cosmic ray nuclei fluxes are expected to be measured with high precision in the near future. High quality data on the antiproton component could give important clues about the nature of the astronomical dark matter. A very good understanding of the different aspects of cosmic ray propagation is therefore necessary.

In this lecture, we will briefly describe a two-zones diffusion model where all the physical effects known to be of some relevance in propagation and diffusion are included. We use cosmic ray nuclei data to give constraints on the diffusion parameters. These results are applied to a new evaluation of the interstellar cosmic antiproton flux. We also study and conservatively quantify all possible sources of uncertainty that may affect that antiproton flux. In particular, uncertainties related to propagation are shown to be underdominant with respect to the ones coming from nuclear physics.

1 Introduction

Understanding the composition and spectral features of cosmic rays has always been an astrophysical challenge. On one hand, the observational data have long been scarce and suffered from large uncertainties. On the other hand, the theoretical predictions to which these data should be compared to have also suffered from several drawbacks. Composition and spectra arise from the nuclear interaction of an initial distribution of energetic particles with interstellar matter (*spallations*) and their electromagnetic interactions with galactic magnetic fields (*acceleration* and *diffusive reacceleration*). First, the nuclear cross sections to be used were not very well known until recently. Second, cosmic rays are sensitive to magnetic field scale inhomogeneities (*diffusion*), which are not well observed. Third, composition and spectra are altered as the cosmic rays enter the solar magnetic field, so that some more modelling has to be done in order to infer interstellar spectra from observations.

In the next Sections, we will outline a diffusion model able to take into account simultaneously all the physical effects considered of some relevance for the propagation of cosmic-rays in the Galaxy and will apply it to calculate the spectra of stable nuclei and antiprotons.

2 The diffusion model

It has been recognized for a long time that the relevant physical propagation model to be used is the diffusion model (Berezinskii & al. 1990, Maurin et al., 2001), though the so-called leaky box model has been widely preferred for decades because of its simplicity. The geometry of the problem used here is a classical cylindrical box whose radial extension is $R=20$ kpc, with a disk of thickness $2h$ ($h=100$ pc) where all the sources are located, and a diffusion halo whose half-height L is an unknown parameter. Diffusion, which occurs throughout disc and halo with the same strength, is independent of space coordinates. The Solar System is located in the galactic disc ($z = 0$) at a centrogalactic distance $R_{\odot} = 8$ kpc.

The steady-state differential density $N^j(E, \vec{r})$ of the nucleus j as a function of energy E and position \vec{r} in the Galaxy, is given by the diffusion equation:

$$\nabla \cdot (K^j \nabla N^j - V_c N^j) - \frac{\partial}{\partial E} \left(\frac{\nabla \cdot V_c}{3} E_k \left(\frac{2m + E_k}{m + E_k} \right) N^j \right) \quad (1)$$

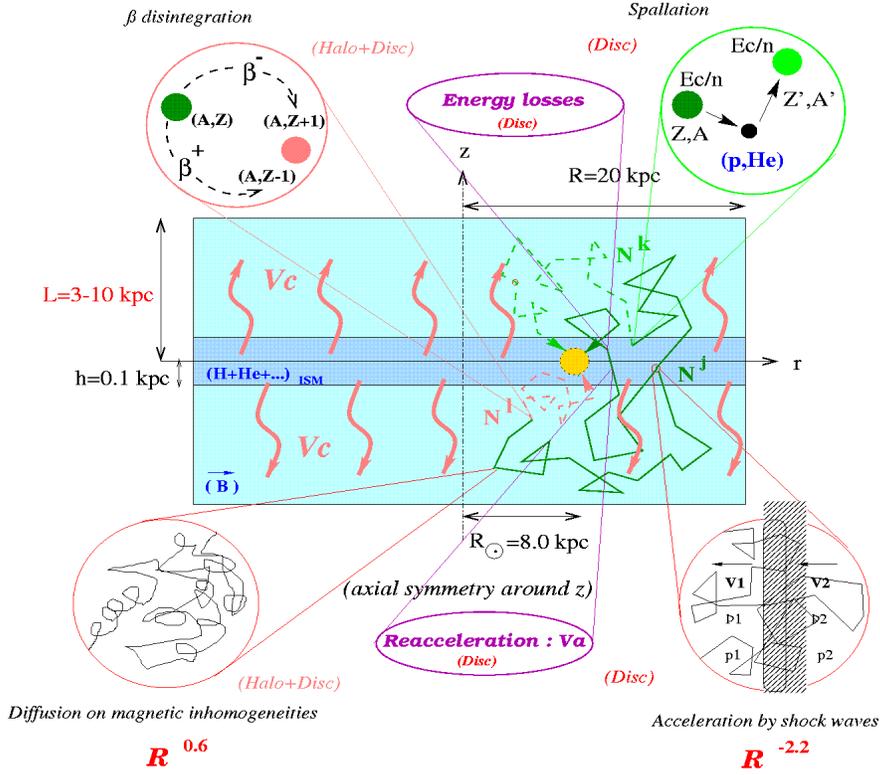
$$+ \frac{\partial}{\partial E}(b^j N^j) - \frac{1}{2} \frac{\partial^2}{\partial E^2}(d^j N^j) + \tilde{\Gamma}^j N^j = q^j + \sum_{m_k > m_j} \tilde{\Gamma}^{kj} N^k$$

The first terms represent diffusion (K^j is the spatial diffusion coefficient; we assume $K = K_0 \mathcal{R}^\delta$, where \mathcal{R} is the particle rigidity) and convection (V_c is the convection velocity). The divergence of this velocity, expressed in the next term gives rise to an energy loss term connected with the adiabatic expansion of cosmic rays. Further, we have to take into account ionization and coulombian losses, plus a reacceleration term in first order derivative (all included in b^j) and finally a second order derivative in E for the associated second order term in reacceleration (d^j is the energy diffusion coefficient). The last term of the l.h.s. takes care of the disappearance of the nucleus j ($\tilde{\Gamma}^j$ for short) due to its collisions with interstellar matter (ISM). In the r.h.s., the source term q^j takes into account the *primary* production and acceleration of nuclei described by an injection spectrum (for the sake of clarity, we have not written down the terms describing the contribution of radioactive species). Finally, the last term is for the *secondary* j sources, namely spallation contribution $\tilde{\Gamma}^{kj}$ from all other heavier nuclei. All the details about the assumptions on the various terms in the diffusion equation may be found in Maurin et al., 2001.

Fig.1 represents a schematic view of the geometry of the Galaxy and the physical effects acting on a cosmic ray.

This equation may be solved analytically using a development over the base of Bessel functions. The solutions must then be treated numerically for the inclusion of energetic losses, effective at low energies: ionization losses over the neutral interstellar matter, Coulomb interactions over completely ionized plasma, dominated by scattering off thermal electrons, and reacceleration (parameterized by means of the Alfvén velocity V_A). With this semi-analytical approach, we obtain the interstellar spectrum for each nuclear species. To compare calculations with observations, one has to take into account the effects of the solar wind on the particles entering the heliosphere. To this aim, usually people employ the so-called force field approximation (Perko 1987).

Figure 1: Schematical picture of the diffusion model.



3 Analysis on stable nuclei

We have compared predictions from our model with the ratio the most sensitive to diffusion parameters: a secondary over a primary nucleus. Up to now, the best measured ratio is B/C (Engelmann et al. 1990). We performed a systematic analysis – indeed the first ever done – varying the relevant parameters K_0 , L , V_c , V_a , and δ of our diffusion model. We obtain a lot of configurations giving a good χ^2 and hence able to reproduce the data, as shown in Fig.2.

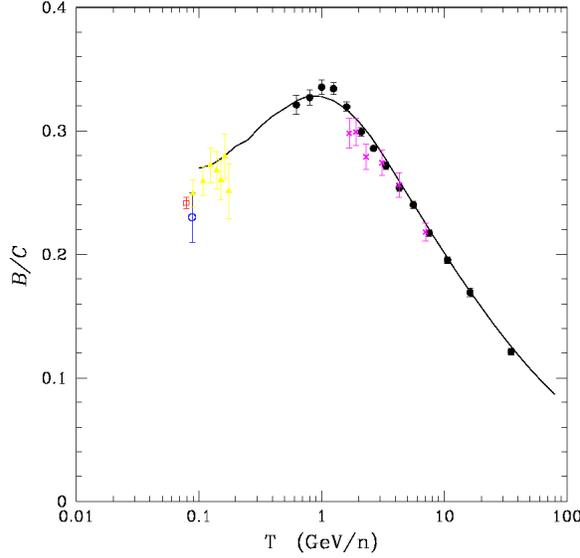


Figure 2: *Computed ratio of $(^{10}\text{B}+^{11}\text{B})/(^{12}\text{C}+^{13}\text{C}+^{14}\text{C})$ for a configuration giving a reduced $\chi_r^2 \approx 1.2$.*

We tested the values for δ permitted by B/C data. As an example, for a halo thickness of 3 kpc, we find that δ is allowed to vary between 0.5 and 0.84. In the whole parameter space, the range of δ extends from approximately 0.45 to 0.85. In particular the value $\delta = 0.33$ corresponding to a Kolmogorov-like turbulence spectrum is strongly disfavoured ($\chi^2 > 100$). For intermediate values of δ , good models are obtained for the full range in L ; for low values of δ , models with a small halo size L are excluded; in particular for $\delta < 0.40$, there is no good model with $L < 25$ kpc. Finally, for high values of δ , models with a large halo L are excluded.

For each $V_a/\sqrt{K_0}$ and V_c we varied L and K_0/L and the best values for the χ^2 are depicted in Fig. 3. The very remarkable result is that we find no model having a good χ^2 without convection ($V_c = 0$) or without reacceleration ($V_a = 0$).

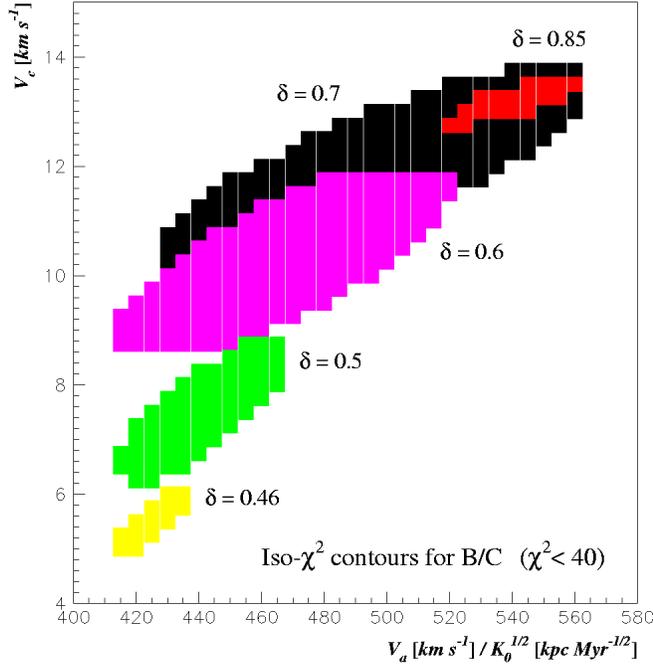


Figure 3: Models with different values of δ are shown. For each value of V_c and $V_a/\sqrt{K_0}$, only the best χ^2 value is retained when the other parameters L and K_0/L are varied. The figure displays the contour levels for $\chi^2 < 40$.

4 Secondary antiproton flux

The study of the cosmic ray antiproton spectrum has been a great challenge since the first measurements made at the end of the seventies. Various primary antiproton sources have been proposed (Silk & Srednicki 1984; Mitsui, Maki, & Orito 1996). The case of supersymmetric sources – relic neutralinos in the dark galactic halo – has received a particular attention and constraints on SUSY parameters have been investigated by comparing experimental data to theoretical predictions (Bottino et al. 1998; Bergström, Edsjö, & Ullio 1999). A major problem with this comparison is an accurate estimation of the background secondary antiproton flux. We will consider it as “background” flux,

having in mind the possibility of using it to determine whether one of the primary components (such as from supersymmetric relic particles or evaporating primordial black holes) could be seen against it or not.

The secondary antiprotons are yielded by the spallation of cosmic ray nuclei (proton and helium) over the interstellar medium. Recent measurements made by the balloon-borne spectrometer BESS and by the AMS detector during the space shuttle flight dramatically reduced the uncertainties both on primary proton and helium spectra. Consequently, the uncertainties on the calculated secondary \bar{p} spectrum due to incoming primaries are negligible.

Whereas p-p interactions are clearly the dominant process for secondary antiproton production in the Galaxy, p-nucleus and nucleus-nucleus collisions should also be taken into account. They not only enhance the antiproton flux as a whole but can change its low energy tail, mostly for kinematical reasons. So we calculated the total antiproton yield considering p-p, p-He, He-p and He-He interactions. Unfortunately, very few experimental data are available on antiproton production cross-sections in nuclear collisions. A model-based evaluation is therefore necessary. Antiproton production *via* the proton-proton interaction was parameterized according to Tan and Ng (1982, 1983). The Monte Carlo program DTUNUC¹ version 2.3 was used to evaluate the cross-sections for p-He, He-p and He-He antiproton production reactions. The resulting cross-sections have been compared with experimental data on proton-nucleus collisions.

Once they have been created, antiprotons may annihilate on interstellar protons. This process dominates at low energy, and its cross-section has been taken from Tan and Ng (1983). Also, antiprotons may survive inelastic scatterings where the target proton is excited to a resonance: these so-called tertiary antiprotons do not annihilate but lose a significant amount of their kinetic energy. We then propagate and solar-modulate the antiproton rate exactly as for stable nuclei. For a complete discussion on the above-discussed interactions and on the treatment of the tertiary component, we refer to Donato et al., 2001.

Fig. 4 displays this computed antiproton flux along with experimental data collected by the BESS experiment (Orito et al. 2000, Maeno et al. 2000) for a reference set of parameters (see Donato et al., 2001). The dotted lines represent the contribution to the total flux coming from the various nuclear reactions.

¹<http://sroesler.home.cern.ch/sroesler/>

First of all, we notice that the calculated spectrum agrees very well with the BESS data points. This strong result gives confidence in our consistent treatment of nuclei and antiproton propagation. Second, even if the main production channel is the spallation of cosmic ray protons over interstellar hydrogen, we see that the contribution of protons over helium is very important, particularly at low energies (where a hypothetical primary signature would be expected).

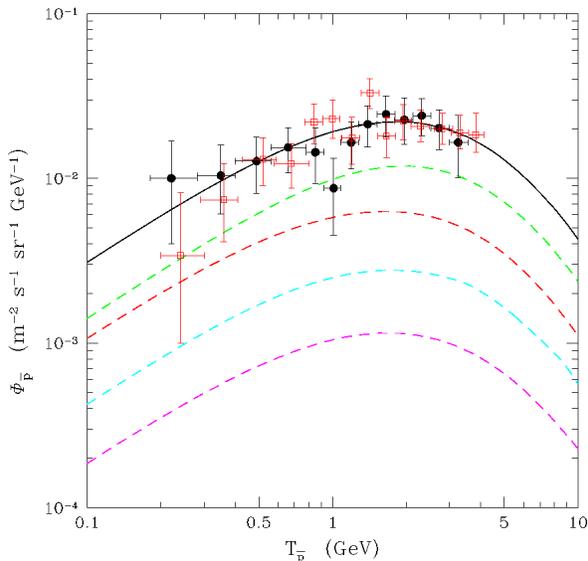


Figure 4: *Solid line shows the total secondary antiproton spectrum for the reference set of diffusion parameters (see text for details). Dashed lines are the contributions to this total flux from various nuclear reactions (from top to bottom: p - p , p - He , He - p and He - He). Data points are taken from BESS 95+97 (filled circles) and from BESS 98 (empty squares).*

Since the propagation parameters are not perfectly known, some uncertainty must affect the antiproton spectrum. To estimate it, we calculated the antiproton spectra corresponding to all the combinations of the free parameters (δ , K_0 , L , V_c and V_a) giving a good fit to B/C. The result is presented in Fig. 5. The two curves represent the minimal and the maximal flux obtained with this set of parameters. The undeterminacy is 9% from 100 MeV to 1 GeV, reaches

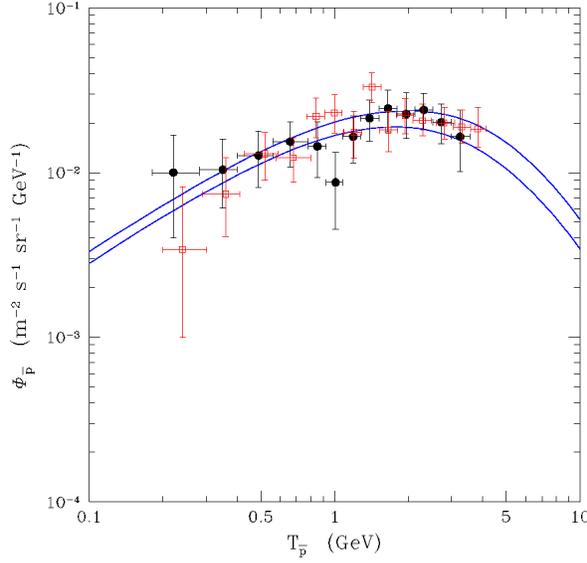


Figure 5: *Antiproton spectra generated with the whole region of parameter space consistent with B/C (Fig. 7 of Paper I). The resulting bounds give an estimation of the uncertainty due to the undeterminacy of the diffusion parameters (data as in the previous figure).*

a maximum of 24% at 10 GeV and decreases to 10% at 100 GeV. It may be considered quite conservative

The uncertainties on the antiproton production cross-sections from p-He, He-p and He-He reactions have been evaluated using the most extensive set of experimental data available (see Donato et al. 2001 for details). All those measurements have been compared with DTUNUC computations. Results on the secondary antiproton spectrum are presented in Fig. 6. The shift of the upper and the lower curve with respect to the central one is of the order of 22–25 % over the whole energy range, dominating uncertainties due to propagation. Many other minor sources of uncertainties have been estimated and discussed in Donato et al.,2001, and they do not add more than few % to the undeterminacy of the flux.

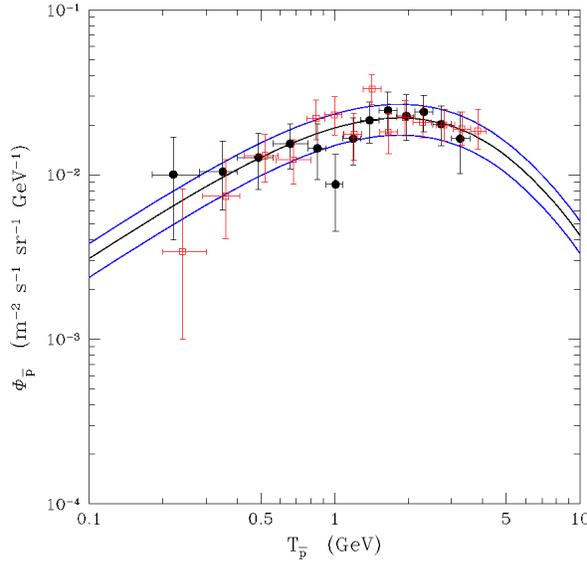


Figure 6: *The antiproton spectrum has been computed with extreme values of DTUNUC nuclear parameters. The central line is the reference curve showed in Fig. 4, while upper and lower curves correspond respectively to the maximum and minimum of the antiproton production rate. These two bounds give an estimation of the uncertainty due to the undeterminacy of the nuclear parameters (data as in the previous figure).*

5 Conclusions and perspectives

We have discussed the propagation of stable nuclei and antiprotons in a two-zone diffusion model, which is shown to reproduce data quite well without any further adjustment. A deep understanding of the physical processes leading to acceleration and propagation of galactic cosmic rays, and more severe constraints on the diffusion parameters, might be got thanks to the several promising space experiments planned for the next decade.

6 Acknowledgements

The author gratefully acknowledges an Istituto Nazionale di Fisica Nucleare postdoctoral fellowship.

References

1. V.S. Berezinskii, S.V. Bulanov, V.A. Dogiel, V.L. Ginzburg, & V.S. Ptuskin, *Astrophysics of Cosmic Rays* (Amsterdam: North-Holland) (1990)
2. L. Bergström, J. Edsjö, & P. Ullio, *Astrophys.J.* 526, 215 (1999)
3. A. Bottino, F. Donato, N. Fornengo, & P. Salati *Phys. Rev.* **D58** 123503 (1998)
4. F. Donato *et al.*, to appear in *Astrophys.J.*, astro-ph/0103150
5. J.J. Engelmann *et al.*, *Astronomy and Astrophys.*, 233, 96 (1990)
6. T. Maeno *et al.*, astro-ph/0010381.
7. D. Maurin, F. Donato, R. Taillet, P. Salati, *Astrophys.J.* **555**, 585 (2001)
8. T. Mitsui, K. Maki, & S. Orito, *Phys. Lett.* **B389** 169 (1996)
9. S. Orito *et al.* *Phys. Rev. Lett.*, 84, 1078, 2000.
10. J.S. Perko, *Astronomy and Astrophys.*, 184, 119, 1987.
11. J. Silk, & M. Srednicki, *Phys. Rev. Lett.*, 53, 624 (1984)
12. L.C. Tan, & L.K. Ng, *Phys.Rev.* **D26** 1179 (1982)
13. L.C. Tan, & L.K. Ng, *J. Phys. G: Nucl.Phys*, 9, 227 (1983)

BALLOON EXPERIMENTS

Piero Spillantini
INFN & University, Firenze, Italy

ABSTRACT

This lesson deals with the balloon borne experiments dedicated to the direct measurement of cosmic rays before their interaction with the terrestrial atmosphere, and the perspectives of the ballooning technique in the next future. I will adopt the following scheme:

- (1) - Direct detection of cosmic rays:
 - (1.1) a short introduction
 - (1.2) experimental peculiarities
- (2) - Ballooning: the technique.
- (3) - Main observation items by balloon borne experiments:
 - elemental composition and spectra;
 - isotopic composition and spectra;

- particle spectra;
 - antiparticle spectra;
 - search for antinuclei;
 - high Z ($>Fe$) nuclei;
- (4) - The future: Ultra Long Duration Balloon (ULDB) flights.

1 Direct detection of cosmic rays.

1.1 A short introduction

The main problems afforded by the studies of cosmic rays are:

- 1. Identification and study of their sources;
- 2. Propagation and acceleration processes in the Galaxy.

Both 1. and 2. reflect the physical characteristics of different physical objects, such as our Sun, super-novae, and extragalactic objects. Therefore the cosmic ray study can be considered part of the Astrophysics, and it is called "Particle Astrophysics". It bridges between Astrophysics and Physics of nuclei and of elementary particles.

In the nuclear physics and particle physics the experiments are generally conducted by using particle beams that, produced and accelerated by accelerators, hit a target. The object of the experiments is the study of the interaction in the target by the detection, identification and parameter measurement of the particles produced in the interaction.

In the particle astrophysics the characteristics of the interactions of the cosmic particles in the Galaxy and in the terrestrial atmosphere are borrowed from the nuclear physics, and the object of the study is the creation and acceleration of the particles in the sources, their injection in space, their propagation in the Galaxy, their possible further acceleration during the propagation and their interaction with the solar magnetic field in entering our solar system: in one word, the object of the study is the cosmic accelerator and its beams.

Cosmic rays can be studied either outside the atmosphere, by satellite borne or balloon borne instruments (direct detection), or on the Earth surface,

by studying the characteristics of the particle showers by them produced in the atmosphere (Extended Air Showers - EAS technique). The most penetrating components (μ and ν) can penetrate the Earth surface, and are therefore studied by underground, or under-ice or underwater experiments. These three detection techniques (direct detection, EAS and under-'Earth' experiments) are the research means of the Experimental Particle Astrophysics (Fig.1)

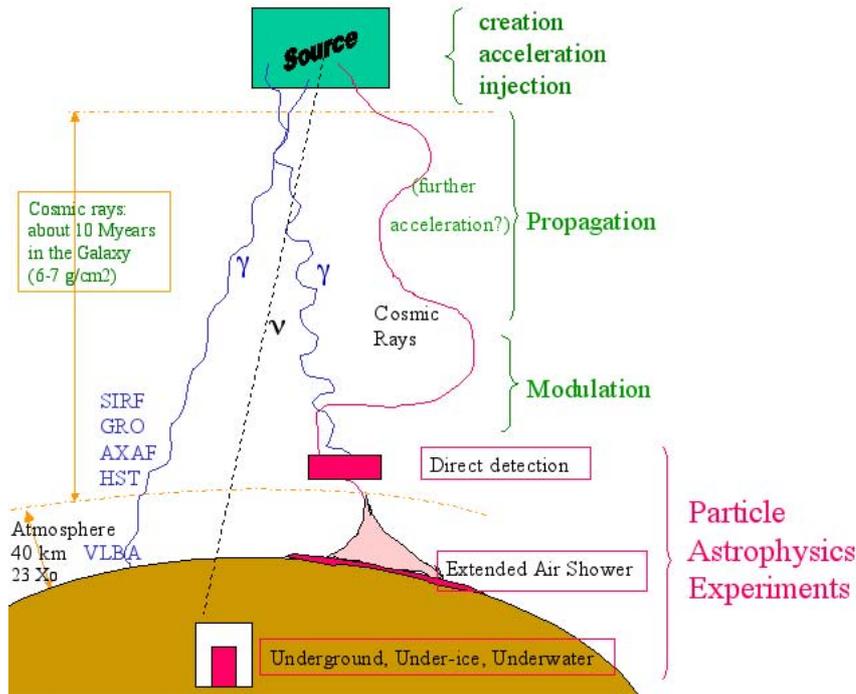


Figure 1: The study of cosmic radiation and the definition of the Particle Astrophysics experiments.

As it is well known, the spectrum of cosmic rays spans many decades of energy, arriving to more than 10^{20} eV, and the flux decreases with energy according to a power law with an exponential index that is roughly the same at all the energies, about -2.7 up to 3×10^{15} eV and -3.0 up to 5×10^{18} eV, and again -2.7 afterwards (Fig.2)

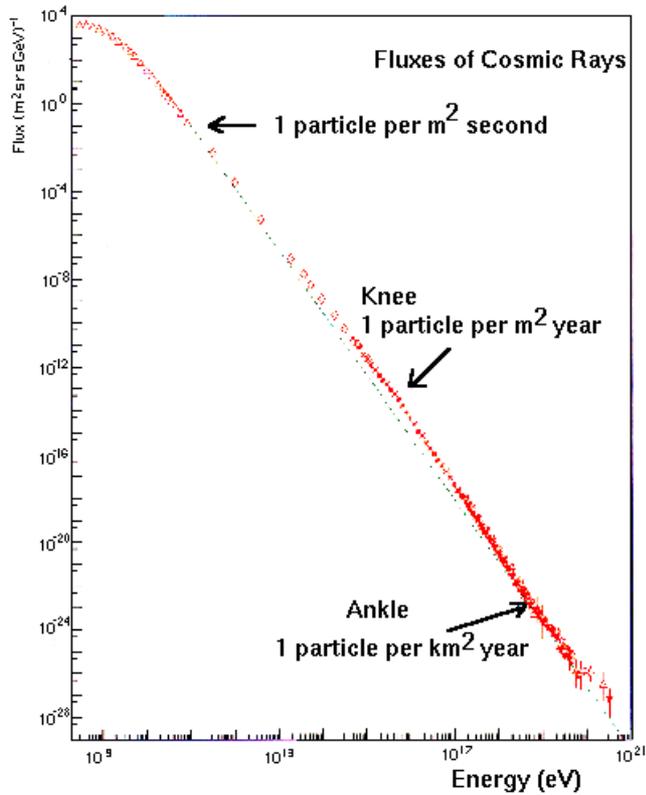


Figure 2: Energy spectrum of cosmic rays as a function of the KE/nucleus

In the energy region where the different components of the cosmic rays can be identified (i.e. up to about 10^{15} eV/nucleus), all the primary components (i.e. those not affected to much by the production of secondary in the interaction of the cosmic particle with the interstellar matter) have the same spectral index of their energy spectrum.

1.2 Primary cosmic rays: experimental peculiarities

There are several peculiarities of the primary cosmic rays that must be taken well in consideration, and in general make somewhat complicated their study:

- The value of the spectral index is very high, between 2.7 and 3.1, what

makes their flux rapidly decrease with increasing the energy. Therefore the measurements of the fluxes of the very high-energy cosmic rays are very difficult, mainly in space (see again fig.2)

- The flux of protons dominates all the other fluxes, so that the study of the other components, specially of the more rare and often more interesting ones, requires complex detection systems. At energies up to some tens TeV, protons constitute about 90% of the total particle flux, and the rest is dominated by helium nuclei by the same proportion, leaving to all the rest (nuclei with $Z \geq 3$, electrons, antiparticles) not more than 2% of the total flux (Fig.3).

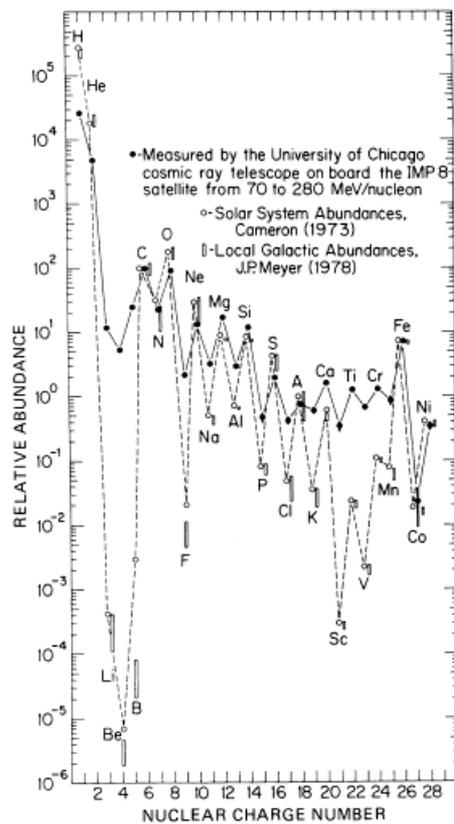


Figure 3: Relative abundance of nuclei in cosmic rays as a function of their nuclear charge

It must at this point be underlined that the detailed knowledge of the spectra of the many isotopes of the cosmic rays is very important for understanding their origin, acceleration, injection processes and diffusion in the Galaxy. Several of them are beta-radioactive (^{10}Be , ^{14}C , ^{26}Al , ^{36}Cl , ^{54}Mn) and can give information on the duration of their storage in the Galaxy. Many others undergo an electron-capture decay (^{41}Ca , ^{44}Ti , ^{49}V , ^{51}Cr , ^{53}Mn , ^{55}Fe , ^{57}Co , ^{56}Ni , ^{59}Ni) and can inform us on the elapsed time between their synthesis inside the stars and their acceleration. Finally there are rare isotopes (^{13}C , ^{18}O , ^{34}S , ^{38}Ar , ^{54}Fe , ^{58}Fe , ^{60}Ni , ^{62}Ni) that can give us information on their synthesis processes. In spite of their extremely tiny flux most of these isotopes can be experimentally observed, provided that it could possibly fly on top of the atmosphere and for enough long time a large acceptance highly selective instrument. It was the case of the LISA instrument selected for the cosmic ray facility ASTROMAG offered about 15 years ago by NASA on board of the FREEDOM Space Station. In fig.4 it is shown the capability that LISA had to have for a deep investigation on the nucleosynthesis, acceleration and diffusion of cosmic rays. When the Space Station FREEDOM was cancelled by USA in 1991, and with it its onboard facility ASTROMAG with its selected experiments (SCINATT, LISA and WIZARD) most of the above mentioned items have been left until now unexplored.

- When created and afterward accelerated, the cosmic ray does not reach us along a straightforward path, but slowly diffuses through the electromagnetic fields filling the Galaxy. The bulk of the cosmic ray flux, around one GeV/nucleon in kinetic energy, reaches us from regions distant up to a few hundreds parsec from the Earth, what means that by their observation we can explore a small fraction of the volume of the Galaxy, less than one percent. The cosmic rays coming from regions near the centre of the Galaxy can prevail at energy exceeding 100 GeV/nucleon.
- For entering the Galaxy and reach the Earth, the galactic cosmic rays must win the solar wind, i.e. the wind of the low energy particles flowing out from the Sun. These flowing particles are charged, and produce a magnetic field whose intensity decreases with the distance from the Sun. Therefore the cosmic ray undergoes the 'bottle effect' that decreases the longitudinal momentum of the particle approaching the Sun. If its momentum is not enough high, at some point the longitudinal component reverses its sense, and the particle keeps

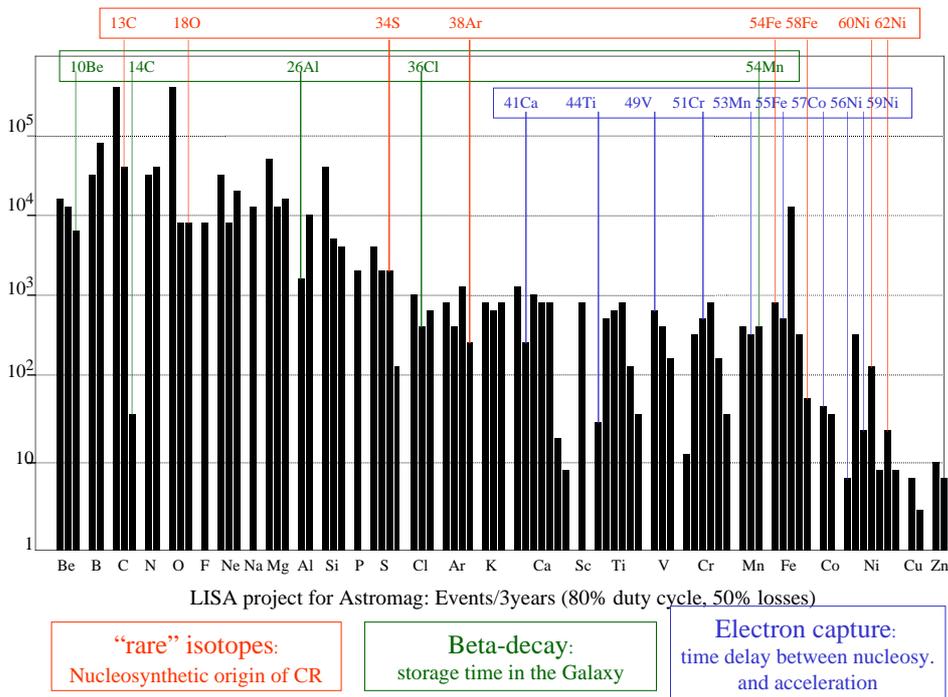


Figure 4: Events/3years foreseen to be observed in the LISA experiment that was selected for the ASTROMAG facility of the cancelled Space Station freedom.

to travel away from the Sun. At the Earth orbit the flux of the low energy galactic cosmic rays is greatly reduced, and their registered energy much lower than that they had in the interstellar space. This phenomenon, known as 'solar modulation', affects all the galactic cosmic rays up to energies of the order of 10 GeV/nucleon. The solar modulation makes necessary to introduce in the fluxes and energy measurements not negligible corrections whose uncertainty prevails on the statistical errors at energies below one GeV/nucleon. During the period of greater activity of the Sun the flow of its wind, and the on it 'frozen' magnetic field, is more irregular, and the effect of these irregularities on the incoming galactic cosmic ray strengths the solar modulation effect. Therefore the flux of the galactic cosmic rays reaching the Earth varies with the 11 years cycle of the solar activity, from a minimum when the Sun is more active to a

maximum, about two times higher, when the Sun is quiet.

- The particle approaching the Earth are swept out by its strong magnetic field. At the equator the incoming particle must have more than 7 GeV/nucleon to reach the terrestrial surface. Since the Earth magnetic field is a dipole, with its axis roughly aligned with the Earth rotation axis (in effect it is inclined of about 11 degrees respect to the rotation axis) and going roughly through its geometrical centre (it passes at about 300 km from it), the strength of the terrestrial magnetic field decreases near the geographical poles, where also low energy particles are admitted to reach the terrestrial surface. For this reason the balloon borne experiments must take in account the latitude at which to perform the experiment, and the satellite borne experiment of the inclination of the orbit of the satellite respect to the equator.

- For the satellite borne experiments, it must also be considered the fact that the dipolar structure of the Earth magnetic field traps most of the particles originated inside its volume, either for interactions or by decay (for example the decay of neutrons). Protons are mainly trapped between one and two terrestrial radii of distance from the Earth surface ('Inner radiation belt') and electrons at more than three terrestrial radii with an intensity still very high at more than six radii ('Outer radiation belt'). These belts are quite wide, but their intensity is reduced in the side of the terrestrial surface due to the interactions of the belt particles with the residual atmosphere still present up to several hundred km of altitude. Satellites must therefore occupy orbits relatively low in altitude (up to 1,000-1,500 km on the terrestrial surface), but not so low, not less than 300-400 km on the terrestrial surface, for not be braked by the terrestrial residual atmosphere. Furthermore, because of the relatively small distance of about 300 km between the magnetic and rotation axes of the Earth, the radiation belts are not centred on the rotation axis of the Earth, so that their distance from the terrestrial surface decreases in the region between the African and South American continents. When the satellites pass over this region (known as 'South Atlantic Anomaly') their electronic devices are subject to an intense radiation and the instruments must be switched off.

- Finally the last, but not least, peculiarity. For the incoming particles the terrestrial atmosphere is very thick, equivalent to 23 radiation lengths for the electromagnetic interactions, 17 collision lengths for the hadronic interactions, and 11 interaction lengths for the inelastic hadronic interactions. In the at-

mosphere the incoming particle interacts and is immediately lost. However, if its energy is enough high, it can produce other particles, and indeed, if also these have enough energy, a shower of particles, where the number of particles progressively increases up to a maximum, afterwards each one of the produced particles has not enough energy to continue the production and is lost.

On the Earth surface only the tail of the showers produced by the very energetic cosmic rays (more than several TeV/particle) can be registered, and for studying the bulk of their shower is necessary to locate very large area arrays of detectors [Extended Air Shower (EAS) detectors] at high altitude on the sea level. There are many of such arrays all around the world. The highest on the sea level are those in Yangbajin (Tibet) and in Chacaltaya (Bolivia), and the largest one AGASA in Japan. A much larger array, covering about $3,000 \text{ km}^2$, is under construction in Argentina, and an analogous one is foreseen in the Utah desert in USA (AUGER project). The EAS detectors supplied us with all the information we have until today on the galactic cosmic rays of energy more than 100 TeV/particle.

With the EAS registration technique we can measure the energy spectrum of the incoming particle, and, registering the number and density of electrons and muons in the shower, we can also distinguish between electromagnetic and hadronic showers, but nothing sure can be said on the kind of hadronic particle that initiated the shower. For identifying the particle it is necessary to operate the experiments on the top of the atmosphere, before the first interaction of the particle with the atmosphere, either in orbit or on board of balloons. Balloons can reach an altitude of about 38-40 km, where the residual atmospheric pressure is less than one percent of the sea level one. As we will see in next paragraph, the duration of balloon experiments cannot exceed, until nowadays, a few days. The energy that the balloon borne experiments can reach depend from the particle to be observed, and is limited by its flux, that decreases by a factor between 50 and 100 for each energy decade. For protons, balloon borne experiments can reach a few TeV. Higher energies, up to a few hundreds TeV/nucleus, are reached in relative long duration repeated flights of the same balloon borne instrument. For more rare components (antiproton and positrons) the maximum reached energy does not exceed 50 GeV, while isotopes cannot be identified at more than a few GeV/nucleon, as it will be later discussed. Until today nearly all the observations of cosmic rays of en-

ergies between about one GeV/particle up to a few hundreds GeV/particle were performed by ballooning, and there were only a few observations made on board of satellites.

Below one GeV/particle the observations by balloon borne experiments become very difficult because of the above-mentioned limitations due to the reduction of the fluxes by the solar modulation and the terrestrial magnetic field. The experiments must indeed be installed on board of satellites or of interplanetary probes, either for profiting of the long duration of the observations or for avoiding the screen due to the terrestrial magnetic field.

2 Ballooning: the technique.

The technique for constructing and operating balloons is not simple. A 'scientific' balloon (i.e. a balloon useful as vehicle for performing scientific experiments) is constituted by a very thin sheet (a few tens of microns of Mylar) that divides two gasses: the gas filling the balloon (usually helium) and the outside atmosphere. For the largest balloons the area of this sheet is a few hundred thousand square meters, and must be immune from holes at the level of a mm^2 . The sheet is subdivided in sectors, glued to strong long strips that must support the services and the instrument, and also all these junctions must be immune from leakage at the same level of a mm^2 along the several km of total length of the junctions. Last difficulty: such big balloons cannot be tested before to be used, and only a few industries can produce them with a reasonable success rate. The so-called 'class A' balloons filled with helium gas reach a volume of 2.8 million of cubic meter at $5g/mm^2$ of residual atmosphere and have a total lifting power of about 11 t, of which 5 t for the material constituting the balloon, 3 t for the services (parachute, etc..) and 3 t dedicated to the instrument.

A balloon-craft is indeed a quite big and complicate machine, difficult to be handled because of the huge dimensions (see in fig.5 the typical dimensions of a 1.0 million cubic meter balloon) and its total mass; only a few crews in the world are able to handle the biggest balloons. The crew which launches the balloons is formed by more than 10 specialists that must spend a few months on the launch field, usually in desert regions (the balloon must obey severe rules for what concerns its trajectory and the landing of the instrument, in order to avoid damages to people and things). Launch operations on the field require

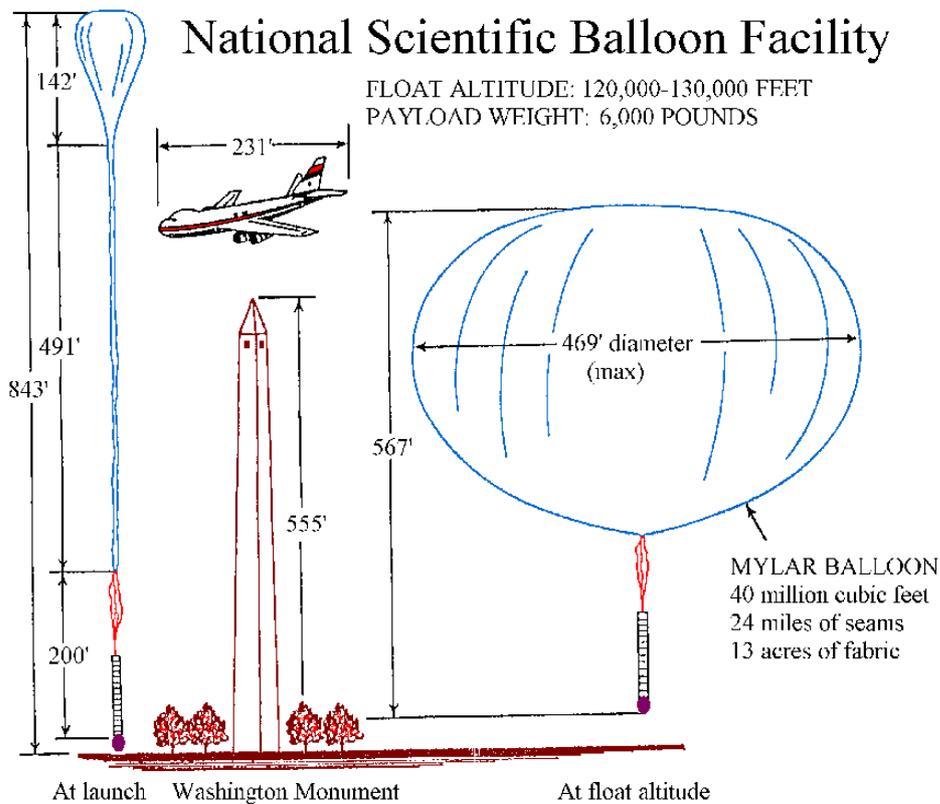


Figure 5: Geometrical dimensions of a 1 Million cubic meter balloon of the National Scientific Balloon Facility of NASA.

the use of a few heavy trucks, either for the bottles of helium gas or for the other services, a special heavy truck for holding the instrument high on the ground before its release (the release is allowed only when the balloon begins to pull perpendicularly to the ground, in order to avoid that the instrument hurts the ground before starting to ascend), an airplane for following the balloon-craft along the flight trajectory and a helicopter to recover the instrument after the flight. Therefore the launch of a big balloon is not a cheap affair, and costs more than half million of dollars, not taking in account the about 200 thousand dollars needed for the balloon.

The sheet of Mylar dividing the helium from the atmosphere is not enough strong for supporting differences of pressure between the two gasses; therefore the balloon is much more similar to a mongolfier than to a usual stratospheric balloon. Its bottom is open in order to leave the helium exit when the helium volume increases due to the increasing of the external temperature. In the night the volume occupied by helium decreases, the balloon shrinks and begins to fall. To recover the altitude a quantity of material (small balls of lead) must be released in order to do the balloon to recover its altitude. The quantity of material to be released (named 'ballast') is about 10% of the mass of the instrument, therefore not negligible, and conflicting with the other massive components to be carried, as for example the batteries supplying the electric power to the instrument. This is one of the reasons why the flight duration of the biggest balloons cannot exceed one or two days. Other limitations to the balloon flight duration are constituted by the speed and direction of the high altitude wind transporting the balloon at its floating altitude. The speed of the wind must be very slow in order to keep the balloon at a distance of not more than 800-1,000 km from the operation site, and speed and direction must be such to guarantee that the balloon does not invade forbidden areas. This put limits on the periods useful for the launch: the crew must wait for inversion of the direction of the high altitude wind, what happens only two times a year and during a short period of a few weeks.

Part of the above limitations can be overcome launching the balloon at very high latitudes in the 'polar summer', in manner that the balloon will be always illuminated by the Sun avoiding the discharging of the ballast each 24 hours. Because of the constant flow of the high altitude wind the balloon makes a circular path, returning to the launch zone (Fig.6).

This is in effect possible, and several balloon borne experiments have been made either in the Antarctica continent or starting from the most northern part of Scandinavia, with duration of 10 or more days. However this kind of launches is very difficult for the ambient conditions, very expensive for the involved resources, and limited to relatively low mass instruments of only a few hundreds kg and not requiring much electric power.

An intense R&D activity is undergoing for producing 'closed' balloons, i.e. balloons that contain the filling helium and support the pressure variations when the Sun's illumination changes. The material constituting the balloon

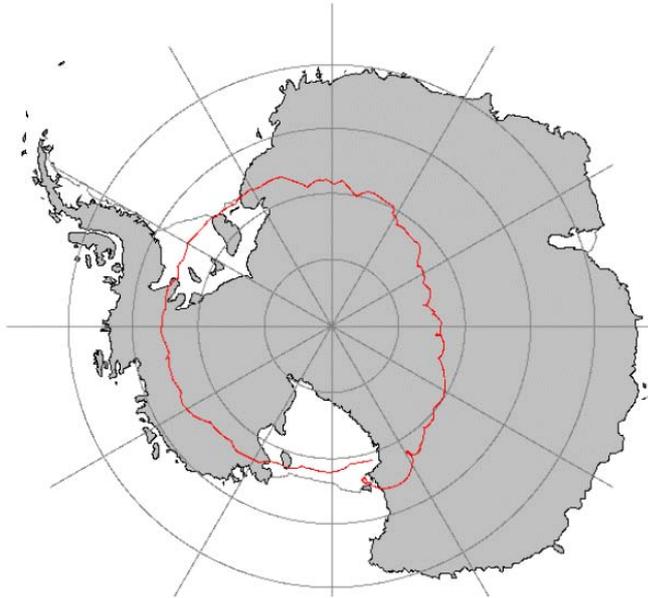


Figure 6: Circular path of the balloon in the Antarctica.

surface is a complex multi-layer, and the requirements concerning the absence of holes is also more stringent than for the open balloons. Such balloons should navigate around the Earth at constant latitude transported by the winds that steadily circulate from West to East, and should stay at the floating altitude for many weeks, hopefully for three or more months [this ballooning program is named "100 days ballooning", or also "Ultra Long Duration Balloon (ULDB) flight" program]. For using such balloon crafts, the instrument must be much more complex than for the nowadays flights, and must include solar batteries for producing the needed electric power, powerful mass memories on board, use communication satellite nets for the control and the data downlink, etc..., and must have many other of the features of a satellite borne instrument.

3 Main observation items by balloon borne experiments.

Before describing the most important observations made by balloon borne experiments, let me give a didactic scheme of the 'jargon' used for the measure-

ments in cosmic rays experiments:

- The flux of particles is usually differential in the kinetic energy of the particle (usually in MeV), the solid angle (in sr) and the area (in cm^2) on which the instrument accepts the particle, and the time (in s).
- The acceptance of the instrument is expressed in cm^2sr and it is called Geometry Factor, and indicated by GF.
- The acceptance multiplied the time gives the 'Exposure', that is indeed expressed in $cm^2 \times sr \times s$.
- The Maximum Detectable Rigidity (MDR) is the rigidity R at which the error ΔR in its measurement is equal to R ; it is expressed in GV/c . For particles with unitary charge the MDR is identical to the Maximum Detectable Momentum (MDM) that therefore results expressed in GeV/c .
- The results of the measurements are usually expressed by the differential flux of the observed particle as a function of its kinetic energy. However the adjective 'kinetic' is often omitted. When convenient the differential flux is integrated on one or more of the differentiating variables, as for example on the GF, or on the exposure, or on the energy. When integrated on the energy we obtain the integral spectrum, giving the total flux of the observed particle integrated from the considered energy up to infinity. At high energies, where the nature of the incident particle cannot be identified, the energy reported in the abscissa is that of the particle, and indicated as energy/particle, when we do not know nothing about the particle, or energy/nucleus when we know that the particle is a hadron, i.e. not an electron or a γ or a ν . At lower energies, when the particle is identified, the fluxes can be reported as a function of the energy/nucleon (usually simplified in energy/n) if the particle is a nucleus, or simply of the energy if it is an elementary particle or a γ or a ν . Often the energy spectra (either differential or integral) are divided by the unity reported in the abscissa (energy, or energy/n, or energy/nucleus or energy/particle) elevated to the index of the energy spectrum in the considered energy interval. This allows putting in evidence in the plots the variation of the spectral index. For special cases (such as the fluxes of rare particles or of antiparticles or for the isotopic composition) the results are expressed by the ratio of two fluxes as a function of the energy at which the ratio is considered.

The ballooning technique is used in many kinds of scientific experiments, not only for measuring cosmic rays. In fact most of the balloon borne scientific

experiments are dedicated to the study of the electromagnetic component of the radiation reaching the Earth, from the mm wavelengths up to the gamma ray wavelengths. A number of experiments are dedicated to the study of the atmosphere, some to the observation of the Earth surface and a few also to the observation of the Sun. The balloon borne experiments dedicated to the study of cosmic rays, i.e. the charged component of the cosmic radiation, are about 20% of the total. A nearly complete list of the most recent balloon borne instruments, some of them still in use, is reported at the Web address:

<http://www.lhea.gsfc.nasa.gov/docs/balloon/instruments.html>

The list of the acronyms of the cosmic ray balloon borne experiments flown in last years or that will soon fly, is reported in fig.7, where the experiments are subdivided according to the main categories of the afforded physics problems. In the following subparagraphs I will briefly comment the flown experiments

Item:	Scientific problem:	Balloon borne experiments:
Chemical composition and Energy spectra of the elements	Test of 'standard model' Calibration of EAS	ATIC, RUNJOB, JACEE, SOKOL, MSU, RICH, CREAM
Isotopical composition	Accel. and propagation	SMLI, IMAX, ISOMAX
Energy spectra of particles	Identification of sources	BETS, BESS, WIZARD EXPT'S
Energy spectra of antiparticles	Antimatter, dark matter	HEAT, BESS, WIZARD EXPT'S
Search for antinuclei	Antimatter	BESS, WIZARD EXPT'S
Fluxes of high Z nuclei (>Fe)	Nucleosynthesis processes Supernovae rate	TIGER
[μ and γ fluxes in the atmosphere]	Atmosphe. ν and γ fluxes	WIZARD EXPT'S, IMAX, BETS

Figure 7: Main items of the study of charged CR by direct detection

and those which will soon take data, subdivided according these categories, mentioning also the experiments operated on board of satellites or of space stations.

3.1 Elemental composition and spectra.

The experimental observation of very high-energy cosmic rays by their 'direct' detection before their interaction with the atmosphere is limited by the rapid attenuation of their flux with the increasing energy. This is the reason why from the very beginning the observations were made by instruments on board of Earth satellites, as soon as they were available. However because of their costs only a few missions could be made, and in the last two decades this study went on by balloon borne experiments. The balloon borne experiments SOKOL, MSU and RUNJOB launch their instruments from the Kamchatka peninsula, at the extreme edge of eastern Siberia, fly for about one week and recover them in the Volga river region. The JACEE experiment is operated in Antarctica, with circum-continental flights that last 10, and sometimes 20 days. In order to maximize the GF the instruments of these experiments include a thin passive calorimeter, where the energy of the shower is measured by evaluating the number of secondary charged particles produced in the first interaction of the incoming particle inside the calorimeter. The contribution of the π^0 's is recovered by allowing their gammas to interact in a high Z material layer and evaluating the number of electrons and positrons behind. The sensitive material of these thin calorimeters is either photographic emulsions or other sensible materials that can be 'etched' and scanned by a microscope.

The study of the energy spectra and of the chemical composition at very high energy is very important. The cosmic rays are assumed to be produced and accelerated in the explosions of supernovae, but this hypothesis alone does not explain the shape of their spectrum at the knee. In fact beyond the knee energy the CR spectrum should drop to zero, much faster than it does. Building up the total spectrum from the spectra measured for each nuclear species should either to test the "standard model" normally assumed for the supernovae explosion mechanism, or to infer new mechanisms of acceleration of the ultra-high energy CR, or the existence of new sites of production. Furthermore it should allow the cross calibration with the EAS experiments based on the Earth surface, that have difficulties to reach (down from the high energy side) the knee energy and can only give a vague indication of the chemical composition of the primary cosmic rays originating the showers. The low statistics in the knee region slowly accumulated by the experiments in the last decades still does not allow to give valuable indications for solving the above questions.

This work is still going on, and should receive a significant buster by the Ultra Long Duration Balloon (ULDB) flight program under development by NASA and by the ACCESS experiment that NASA propose to install in future on board of the International Space Station (ISS).

3.2 Isotopic composition and spectra.

The instruments for the measurement of the spectra of isotopes are somewhat complex for the need of identifying the isotope. The particle momentum must be measured with a good precision, at the level of a few %, and the velocity measured by adequate Time of Flight and Cherenkov systems. It is therefore difficult to have good performance of the instrument at energies above one GeV/n. The results until now obtained are below one GeV/n, with a few points at a few (2-3) GeV/n for the lighter isotopes. Unfortunately the models of the distribution of the interstellar matter and of the sources in the Galaxy give isotopic spectra very different at energies significantly higher than 1 GeV/n.

The balloon borne instrument ISOMAX was designed and developed for reaching several GeV/n for the most interesting radioactive isotopes: it had a powerful and precise superconducting magnetic spectrometer, an optimised time of flight measurement and several cherenkov counters, and an enough large GF. Unfortunately this instrument was destroyed by wrong manoeuvres during the re-entry from the last flight this summer. No other new instruments are planned for the next future dedicated to the isotopic measurements at enough high energies, and, as above noted in paragraph 1.2, this field will stay for long time unexplored by dedicated experiments.

However for the measurement of the lightest isotopes a significant contribution should come from two instruments to be shortly launched in orbit: the experiment PAMELA of the WIZARD collaboration that will be operated for three years on board of a Russian satellite starting at the end of 2002, and the experiment AMS-02 that will be installed on board of the ISS from 2004 and will also take data for three years. These experiments are optimised for studying the antiparticle components in the primary cosmic rays (see the following paragraph and the lessons dedicated to them), and the many detectors they must use for the antiparticle identification will allow to measure the lightest isotopes up to a few GeV with a good statistics and a low systematic error.

3.3 Particle and antiparticle spectra and search for antinuclei.

Most of the above mentioned experiments gave results concerning the isotopic composition as a by-product of the measurement of the CR antiparticle components. The subject of reconciling the symmetry of the Universe in its content of particles and antiparticles as it should result from the standard cosmology and the presence of only particles in our corner of the Universe moved the cosmologists for several decades. Shortly after the discovery of the CP violation in the weak interactions in '64, Sakharov formulated three hypothesis that were assumed to be a reasonable starting point for explain the apparent contradiction between the fundamental laws of nature and the observations (see the lesson "The antimatter component of cosmic rays and the PAMELA experiment" of this course). Several balloon borne experiments were dedicated to the search for antiparticles and antinuclei, and in the 70's the teams of B.Golden in USA and of E.Bogomolov in Russia identified the first antiprotons in cosmic rays (the positrons were discovered more than 40 years before by Anderson). Then the antiproton spectrum was intensively studied for searching for signals exceeding the background of the antiprotons produced in the interactions of CR's with the interstellar matter. The flux of this background is less then 10^{-6} of the proton flux at 200 MeV, reaches 2×10^{-4} at 10 GeV, afterwards decreases to about 10^{-5} at 1 TeV and continues to decrease at higher energies. Because of these small ratios of the antiproton to proton fluxes, the identification of the antiprotons requires the use of a magnetic spectrometer for selecting the negatively charged particles, leaving antiprotons to be separated from electrons, what is much easier, also because the electrons are much less abundant than protons. Only in a very old experiment of Buffington at very low energies the identification of the antiprotons was obtained by the pattern of their annihilation in an imaging calorimeter without the help of a magnetic field. This experiment gave a result contradicting all the following measurements, probably for the weakness in the identification of the antiprotons without the help of a magnetic spectrometer. The present situation for the measurement of the antiproton on proton ratio is reported in fig.8.

Because of the needed strength of the magnetic field the minimization of the matter on the path of the particle, and the time of flight and other velocity measuring instruments for separating the antiprotons from the electrons, (and in most of the experiments also an imaging calorimeter where to do the

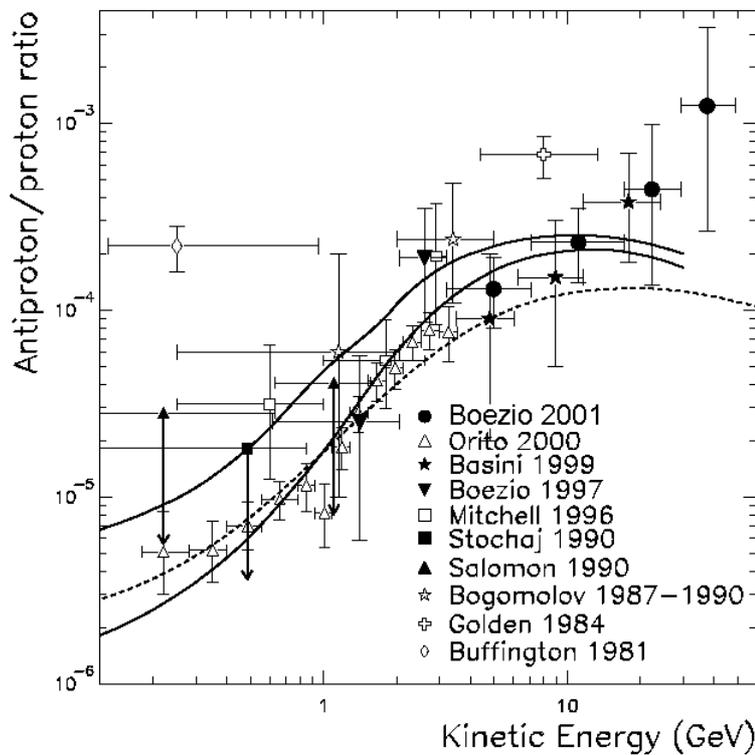


Figure 8: Antiproton/proton ratio: experimental situation, compared with the foreseen secondary production in the interstellar medium ¹⁾.

antiproton annihilate and be recognized), all the experiments for studying the antiproton flux have an enough rich instrumentation for giving useful results in many other channels, such as the measurement of the flux of the positrons (Fig.9), of the lightest isotopes (as above underlined), of the spectra of the nuclei (normally confined to the lightest ones because of the limited abundances of the others), of the fluxes of the different products of the interaction of the primary particle with the atmosphere (registered by someone of these experiments during the ascension phase from the launch site to the floating altitude), and of the spectra of protons and helium nuclei up to energies of 1 TeV/n (which can be performed up to so high energies because of their high fluxes). Finally in fig.10 is reported the situation for the limits reached in the

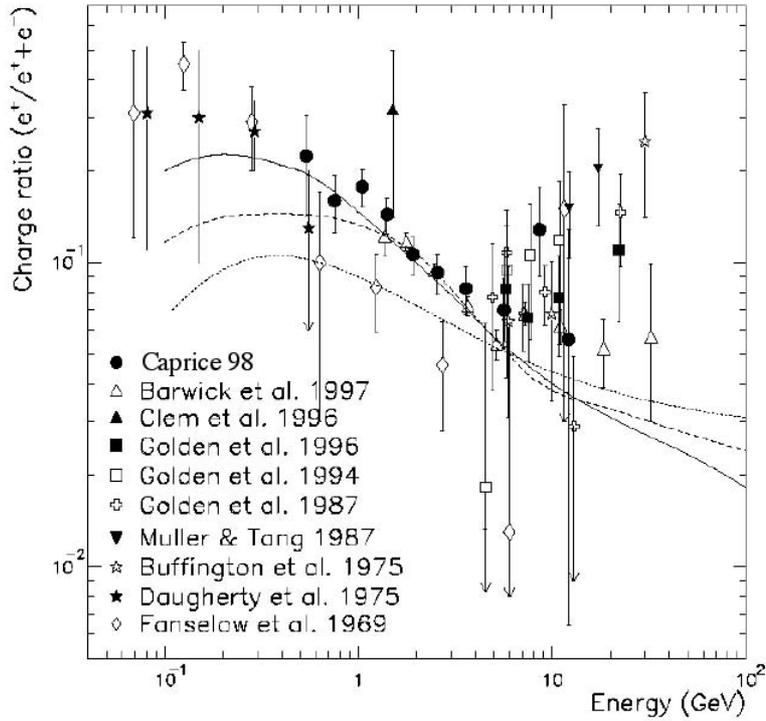


Figure 9: Positron/(positron+electron) ratio: experimental situation, compared with the foreseen secondary production in the interstellar medium ²⁾.

search for antinuclei. In the study of positron and antiproton spectra, balloon borne experiments are limited to relatively low energies (≤ 50 GeV) by the tiny fluxes. However they cannot profit neither of the Long Duration Balloon flights already regularly operated in the Antarctica continent by other experiments, nor of the future ULDB flight program. There are two difficulties. The first one is instrumental: the mass of the instrument (always exceeding 1 t) does not allow the use of such flight occasions. The second one is due to the production of secondary positrons and antiprotons in the residual atmosphere on top of the instrument, that cannot be less than $4 - 5 g/cm^2$. The reason for this limit is the temperature profile of the atmosphere: starting from the terrestrial surface the temperature progressively decreases with the altitude reaching about

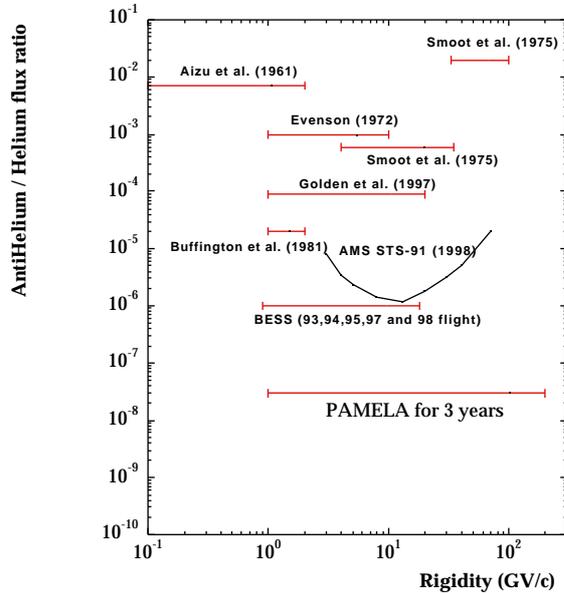


Figure 10: Antinuclei: experimental situation for the antihelium search. In the figure is reported also the sensitivity foreseen for the PAMELA experiment. The AMS-02 experiment foresees to reach a sensitivity of about 10^{-10} , i.e. two order of magnitude lower than that of PAMELA.

$-80^{\circ}C$ at about 20 km, afterwards it increases again for reaching about $0^{\circ}C$ at 40 km, that is the altitude reached by the balloons used for the study of the antiparticle component. At altitudes higher than 40 km the temperature keeps again diminishing, preventing the balloon to further rise because of the shrinking of its volume with the temperature.

Also for this the working group appointed about twenty years ago by NASA for constructing a program for cosmic ray research for the years 1985-1995, advised that these experiments should be performed in orbit. For this research it was selected the above mentioned WIZARD experiment to be conducted on board of the FREEDOM Space Station. When the FREEDOM

Space Station was cancelled, the WIZARD collaboration formed at that time for this experiment began to operate to bring anyway it in orbit. After several launches of balloon borne experiments, the collaboration afforded the construction of the PAMELA experiment. The PAMELA instrument is now under construction and will be launched at the end of 2002 from the Baikonur cosmodrome on board of the REUSURS-DK1 Russian satellite. It will measure the antiproton spectrum up to 190 GeV and the positron spectrum up to 270 GeV, in order to search for signals exceeding the background of the secondary production. Because PAMELA is equipped by several sophisticated detectors, it will give many other by-product results in the channels mentioned in the previous paragraphs. A bigger instrument, the AMS-02 spectrometer, will be installed in 2004 on board of the ISS. It has a much larger GF and the same detector set than PAMELA. The main goal of the AMS-02 experiment is to hunt for antinuclei (until now there are only upper limits for their flux), but it will also measure the antiproton and positron fluxes in an energy range similar to that of PAMELA and give similar by-product results.

3.4 High Z ($>Fe$) nuclei.

The search for the high Z nuclei beyond iron (which is the last element that can be synthesized inside the stars) is important for understanding the violent astrophysical phenomena where these nuclei are synthesized. However their study is extremely difficult, either for the technical difficulty of measuring the value of very high electrical charges, or for the extremely tiny fluxes of these ions at any energy, several order of magnitude lower than those of the iron nuclei. The most important group to be studied is that of the actinides, because their flux and their distribution in charge allow to determine the rate of the 'fast processes' (i.e. of the supernovae explosions) in the Galaxy. The instrumental difficulty has been solved in these last years by using suitably doped glasses that after their exposition to the cosmic radiation can be etched by suitable solvents. The going through heavy ions damage the material and the solvent produces in the material holes whose diameter and depth depend from the charge and the velocity of the going through particle. This technique is somewhat delicate, and has been set up in the last experiments on board of the Long Duration Exposure Facility (LDEF, recovered from space by the Shuttle vehicle after six year of exposition in the space) and of the MIR Space Station. The need

of recovering the sensible material for the etching and the subsequent analysis requires that such kind of experiments be located on board of a visitable space station (as it was the MIR and it is now the ISS) or of a spacecraft that after the exposition in orbit can be reached by the Shuttle and brought back to ground (as it was the case of the LDEF). In principle these kind of research could be made on board of balloons, because the instrument can be recovered after the flight. However short balloon flights are not very useful because of the too small exposure, and the only possibility is to expose large GF instruments in ULDB flights. In fact the experiment selected by NASA for the first demonstration flight of the ULDB program is the TIGER experiment that will be launched in 2003 and will study the elemental CR composition beyond the iron up to calcium. To go beyond calcium also ULSB flights are not enough. For a less near future NASA foresees to put in orbit two experiments that will fly together as a unique structure: ENTICE, an electronic experiment that will study the nuclei beyond the iron up to uranium with a GF of a several $m^2 sr$, and ECCO, based on the above mentioned doped glasses, with an acceptance of several $10m^2 sr$ for measuring also the actinides. After several year of exposition in space of the ENTICE+ECCO complex the sensitive glasses will be retrieved by the Shuttle spacecraft to be etched and analysed on ground.

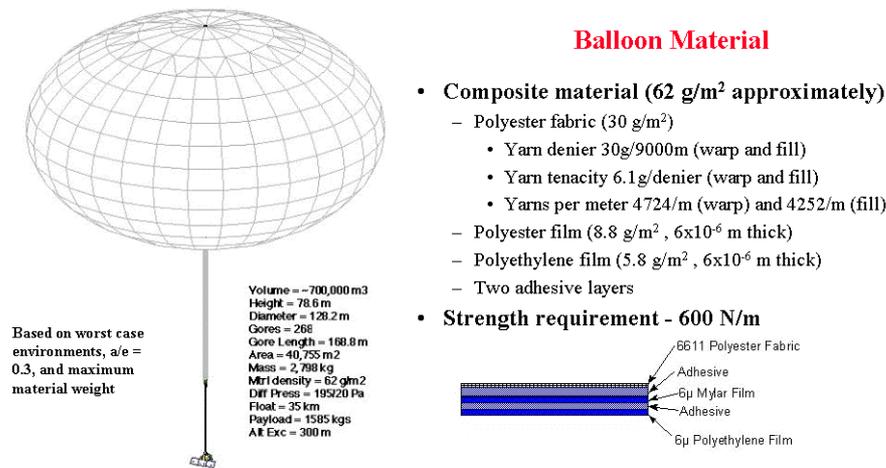


Figure 11: The characteristics of the ULDB for the DEMO experiment TIGER

4 The future: Ultra Long Duration Balloon (ULDB) flights.

The above mentioned ULDB experiment TIGER will be a 'demo' experiment, first of a series that NASA has already in its program for the future, with a cadence of two launch per year. The characteristics of the balloon for the 'demo' experiment are reported in fig.11. NASA therefore is not abandoning the ballooning technique for future scientific experiments. Many of the collaborations conducting the above mentioned balloon borne experiments are presenting proposals for going on in their programs flying their instrument in the framework of the ULDB program. Proposals have also been presented for the large area IR telescope PRONAUS, the large solar telescope FLARE GENESIS, an extra-solar planets finder, all sky surveys for hard X-rays, γ 's (HIREGS), Cosmic Microwave Background (BOOMERANG). In general the balloons flown in the framework of the ULDB program will be used by physicists for atmospheric, astronomic and astrophysics researches, by the USA DoD for weapons and other surveillance systems and by other organizations for commercial services such as telecommunications and advertising.

It must be underlined that the ULDB program is supported by NASA also for its technical fallout perspectives beyond the traditional ballooning activities. In fact pressurized balloons will be used by NASA either in future planetary exploration missions for exploring planets and satellites that have an atmosphere, or for inflatable applications (such as huge trusses, space antennas, huge optical systems, etc...), or for constructing steered 'strato-stations' that flying to the very high latitude of 30-40 km could be used for experiments, Earth observation and communication services.

It is like to say as a conclusion that, wherever an atmosphere exists, the ballooning activity will not die.

References

1. D. Bergström et al. 2000 *ApJ Letters*, **534**, L177 and references therein.
2. M. Boezio et al. 2000, *ApJ*, **532**, 653 and references therein.

ANALYSIS OF COSMIC-RAY DATA: EXAMPLES WITH ANTIPROTONS AND POSITRONS

Mirko Boezio

Sezione INFN di Trieste, Via A. Valerio 2, I-34147 Trieste, Italy

ABSTRACT

Data analysis is a relevant part in the experimental-physics work. Here we will discuss it referring to the analysis of antiprotons and positrons in the cosmic radiation.

1 Introduction

Measurements of cosmic-ray antiparticles (antiprotons and positrons) are important for understanding several aspects of cosmic-ray physics. Their study permits to investigate the origin and propagation of cosmic rays in the Galaxy. Moreover, questions such as: do antimatter exist domains in the Universe? do mini-black holes exist and evaporates? is dark matter made up of supersymmetric particles? can be answered by these measurements.

To address these scientific topics a precise measurement of the antiproton and positron energy spectra in a wide energy range is needed. Fig. 1 shows the existing antiproton flux data. The experimental data are presented along

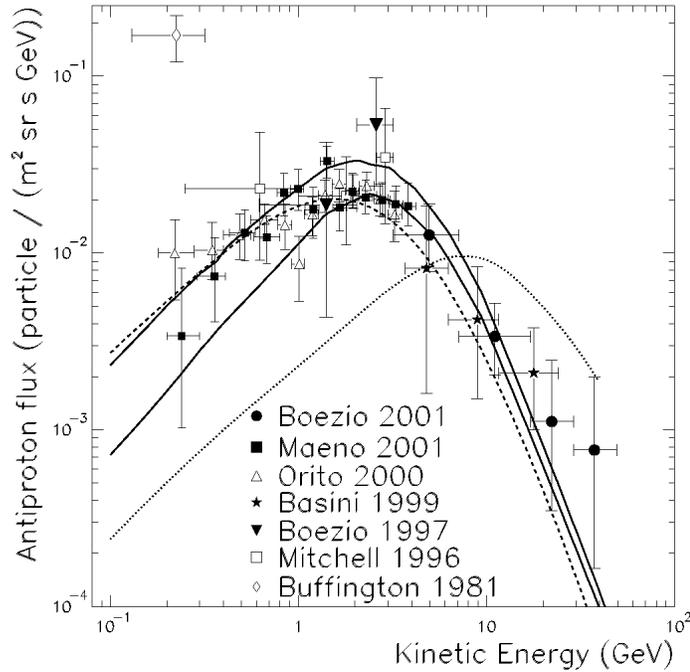


Figure 1: *The experimental \bar{p} spectrum 1, 2, 3, 4, 5, 6, 7) along with theoretical calculations 8, 9, 10) (see text).*

with different theoretical calculations which account for a pure secondary component 8, 9) (all antiprotons are produced by interaction of the cosmic-rays with the interstellar matter, solid and dashed lines) and for a primary component 10) (dotted line). The consistency among most of the experimental data is good and the measurements favour a pure secondary component. However, because of the limited statistics of these experiments, a primary component cannot be ruled out and the data are not precise enough to discriminate be-

tween different propagation models. In the following a brief description of the experimental and analysis procedure is provided.

2 Experimental measurements

Experimental study of primary cosmic rays requires that the measurements are performed outside the Earth's atmosphere or very close to the top of it so to minimize the atmospheric overburden. For this reasons, the experiments are attached to stratospheric balloons or on board of satellites. With the exception of the AMS experiment ¹¹⁾ on the space shuttle, all the existing measurements of antiparticles were obtained by balloon-borne experiments.

The difficulties of these measurements are mostly related to the fact that antiparticles are a rare component of the cosmic radiation. Antiprotons and positrons must be distinguished from a large background of protons and electrons. Furthermore, for balloon-borne experiments there is an additional background of products, such as muons and pions, of cosmic-ray interactions with the overlying atmosphere. Moreover, the data must be corrected for the secondary production in the atmosphere, which becomes increasingly important for energies below about a few GeV. The main goal of these experiments is to provide measurements of the fluxes of the cosmic-rays hence precise determination of the energy of the particle is required. In most experiments this was done using magnetic spectrometers with which the rigidity¹ of the particles was measured. While very precise, this technique, however, requires the study of the rigidity dependent background of spillover protons especially relevant for the antiproton analysis. In fact, due to the limited spectrometer resolution, high rigidity protons can mimic negative tracks and, hence, be tagged as antiprotons. Furthermore, reliable energy spectra measurements requires precise determination of the efficiencies of the various detectors.

Figure 2 shows CAPRICE98 apparatus ¹²⁾, a balloon-borne experiment that gave results both on antiprotons and positrons, that illustrates some typical features of these type of cosmic-ray experiments. The CAPRICE98 apparatus, from top to bottom, consisted of: a gas Ring Imaging Cherenkov (RICH) detector ¹³⁾, a time-of-flight (ToF) system, a magnet spectrometer and a silicon-tungsten imaging calorimeter ¹⁴⁾. The ToF system was used for

¹Rigidity= momentum \times speed of light/charge.

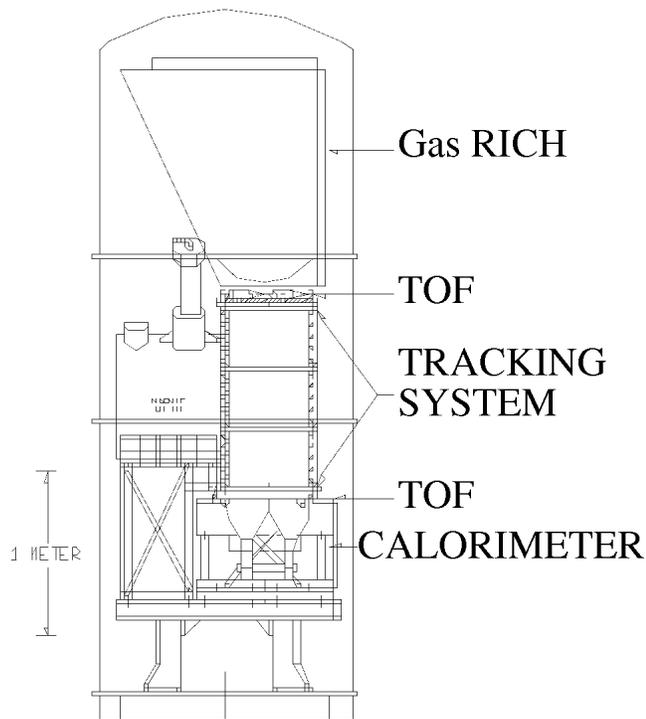


Figure 2: *The CAPRICE apparatus in the 1998 configuration (CAPRICE98).*

the trigger and to measure the time-of-flight and ionization losses of the particles. With this information single-charge down-going particles were selected rejecting helium and heavier nuclei, which are a consistent fraction of the cosmic radiation, and albedo particles, which are secondary particles produced in the atmosphere with an up-going direction. The magnet spectrometer had a superconducting magnet and a tracking device from which information the rigidity and charge sign of the particles were determined. The final selection of \bar{p} in a background of e^- , μ^- and π^- and of e^+ in a background of p , μ^+ and π^+ was performed with the gas-RICH and the calorimeter. The RICH, for particles with velocity above its Lorentz threshold (about 19 for the C_4F_{10}

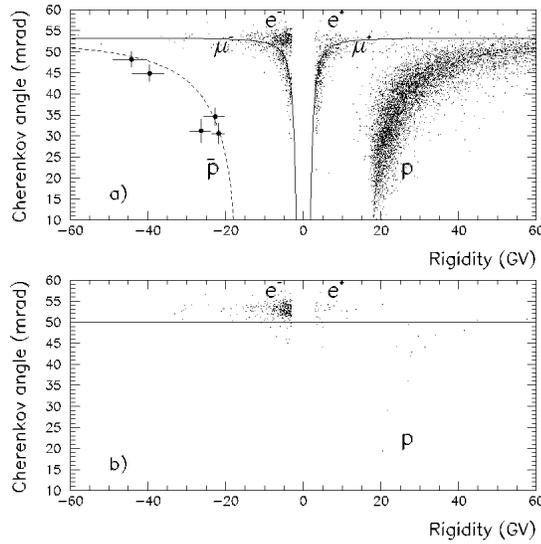


Figure 3: *The measured Cherenkov angle by the CAPRICE98 experiment as a function of rigidity. In a) the solid and dashed lines represent the theoretical values of the Cherenkov angle for muons and (anti)proton respectively. On the negative side, the location of five antiprotons are indicated with black squares together with one standard deviation errors on the measured rigidities and Cherenkov angles. In b) the events surviving the calorimeter electron selection are shown. The solid line indicates the lower limit for the RICH electron selection.*

gas), provided a measurements of their velocity through the reconstruction of the Cherenkov angle ¹³). Comparing this information with the rigidity measured by the tracking system the particles can be identified. Fig. 3 shows the Cherenkov angle as measured by the RICH as a function of rigidity. Fig. 3 a) shows how the RICH information can be used to identify antiprotons: in correspondence of the antiproton band (dashed line) five antiprotons are clearly distinguishable from the e^- and μ^- background. Here they are shown with one standard deviation error bars for both the rigidity and Cherenkov angles measurements. Below the antiproton gas-RICH threshold (about 18 GV) an-

tiprotons can still be identified since no Cherenkov light is detected contrary to muons and electrons. This is valid down to the muon threshold, that is 2 GV. To extend the antiproton identification to lower energies a RICH detector can still be used replacing the gas with a radiator of higher refractive index. For example, a similar configuration of CAPRICE98 apparatus was flown in 1994, therefore called CAPRICE94, where sodium-fluoride (NaF, refractive index $\simeq 1.39$ and threshold Lorentz factor $\simeq 1.4$) was used as radiator and antiproton were identified between 0.6 and 3.2 GeV of kinetic energy ⁵⁾.

It is clear from fig. 3 that the RICH can be used to identify positrons as well. However, in this case, the background of protons is so vast that additional information is needed for the separation. Furthermore, as can be seen in the figure, the RICH cannot separate μ^+ from e^+ above about 4 GV. The additional rejection is provided by the calorimeter. This calorimeter is a sampling calorimeter of 7 radiation lengths in which the electrons shower producing the typical electromagnetic cascades. Thanks to a fine longitudinal and transversal segmentation, along with the measurement of the energy losses in the sensitive elements (silicon layers), electromagnetic showers can be identified by imposing conditions on parameters describing their shape and energy distributions (for more information see ¹⁵⁾). The effect of the calorimeter selection can be seen in fig. 3 b) where only the events of fig. 3 a) passing the electron calorimeter selection are shown. Comparing the two figures it is evident that most of the contamination has been rejected and the electron and positron signals is easily extracted from the remaining background using the RICH information (i.e. selecting all the events above the solid line in figure).

Antiprotons and positrons have also been identified with other detectors. The BESS experiment ³⁾ used a time-of-flight system providing high-resolution velocity measurements with which they selected antiprotons up to about 4 GV. The HEAT-pbar experiment ¹⁶⁾ derived the particle velocity from multiple dE/dx measurements and measured antiprotons up to 50 GV. TS93 ¹⁷⁾ and HEAT ¹⁸⁾ experiments identified e^+ up to 50 GeV combing a transition radiation detector with an electromagnetic calorimeter.

3 Flux calculation

The flux is obtained from the relation

$$\text{Flux}(E) = \frac{1}{T_{live} \times G \times \Delta E \times TF} \times N^{TOA}(E),$$

where T_{live} is the total time during which the apparatus was able to record data, G is geometrical factor, ΔE the width of the energy interval² in which the events were selected and N^{TOA} the number of selected events scaled to the top of the atmosphere. T_{live} is usually obtained scaling the time of data taking by the dead time of the experiment. The geometrical factor accounts for the acceptance of the apparatus and the simplest way of calculating it is with Monte Carlo techniques²⁰). ΔE is the energy bin corrected for ionization losses from the top of the atmosphere to the tracking system where the measurement is performed.

Using the information given by the various detectors described in the previous sections, \bar{p} and e^+ are identified from the recorded data. In this process also the contamination by other particles is estimated and usually either subtracted from the selected sample or the rigidity interval, in which particles are selected, is chosen to have a negligible surviving contamination. Then, to obtain N^{TOA} , the numbers of particles selected in the spectrometer must be corrected for :

1. selection efficiencies;
2. losses of particles because of interaction with the detectors above the tracking system and the atmosphere overlying the apparatus and
3. secondaries produced in the atmosphere.

These corrections are rigidity dependent and have different effects in the flux estimation that are, often, quite significant. Efficiencies are usually estimated using either simulation or the flight data themselves. Both methods require a careful study of possible biases and variation in the performances of the detectors during the flight. For example, pressure changes, which normally happen during a balloon flight, for a gas-RICH as the CAPRICE98 RICH result

²Usually cosmic-ray fluxes are expressed in number of particles per (m² sr s GeV) and the energy is the kinetic energy.

in variation of the refractive index of the radiator and, consequently, of the Cherenkov threshold and number of Cherenkov photons produced. This clearly affects the performance of the detector and its selection efficiency. Overlooking these effects can result in significant errors on the fluxes. Sometime is not possible to precisely determine these effects hence, the correct way to proceed is to estimate the uncertainty on the knowledge of these effects and propagate it as a systematic uncertainty in the flux calculation.

Atmospheric secondaries are obtained from calculations and they represent a significant component of the measured antiprotons and positrons. For example fig. 4 shows the e^- and e^+ spectra measured by the CAPRICE94 experiment ¹⁵⁾ at 3.9 g/cm^2 of residual atmosphere. The solid and dashed lines show the estimated contributions of atmospheric e^- and e^+ , respectively. The data were taken at a vertical geomagnetic cutoff of about 0.5 GV. It can be seen that, except below 1 GeV, the secondary electron component accounts for a few percent of the primary component, hence, assuming that all electrons are of primary origin would result in a small error. However, this does not hold for positrons: for all energies the secondary component is an important fraction of the positron flux and the data must be corrected for it. This also means that uncertainties on the secondary calculation will reflect significantly on the flux estimations.

4 Conclusions

Measurements of cosmic rays are very challenging because of the environmental difficulties in which these experiments take place. This, consequently, affects the data analysis and it is even more true for antiparticle searches because of the paucity of these particles in the cosmic radiation. Therefore, along with the study of the detector outputs to derive good selection conditions and to estimate the background contamination a careful analysis of all assumptions made for estimating the final result has to be performed. From these analysis statistical and systematic errors are derived and they have to be included in the final result. For example, the data in fig. 1 include both statistical and systematic uncertainties. However, it is worth pointing out that an additional systematic uncertainty is not clearly indicated in this figure and often overlooked: the uncertainty on the energy (rigidity) determination. This quantity is often not very well known since it requires an independent energy measure-

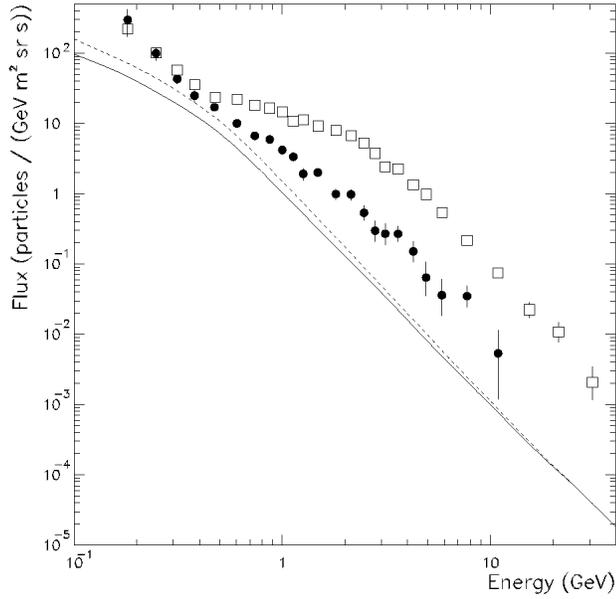


Figure 4: *The electron (□) and positron (●) spectra measured at top of the payload by the CAPRICE94 experiment ¹⁹⁾ and the calculated atmospheric electrons (solid line) and positrons (dashed line) at 3.9 g/cm² of residual atmosphere ²¹⁾.*

ment that is often not available during the experiment. Also calibration at a test beam facility at ground prior or after the flight is often not sufficient to represent the experimental conditions since the devices can suffer variation and misalignments during the launch phase.

References

1. M. Boezio *et al.*, astro-ph/0103513 (2001).
2. T. Maeno *et al.*, Astropart. Phys. **16**, 121 (2001).
3. S. Orito *et al.*, Phys. Rev. Lett. **84**, 1078 (2000).

4. G. Basini *et al.*, in Proc. 26th Int. Cosmic-Ray Conf., **4**, 77 (The University of Utah, Salt Lake City, 1999); see also M. Hof *et al.*, *Astrophys. J.* **467**, L33 (1996).
5. M. Boezio *et al.*, *Astrophys. J.* **487**, 415 (1997).
6. J. Mitchell *et al.*, *Phys. Rev. Lett.* **76**, 3057 (1996)
7. A. Buffington, S. M. Schindler and C. R. Pennypacker, *Astrophys. J.* **248**, 1179 (1981).
8. M. Simon, A. Molnar and S. Roesler, *Astrophys. J.* **499**, 250 (1998).
9. L. Bergström, J. Edsjö and P. Ullio, *Astrophys. J.* **526**, 215 (1999) and L. Bergström and P. Ullio, private communication (1999).
10. P. Ullio, astro-ph/9904086 (1999).
11. J. Alcaraz *et al.*, *Phys. Lett. B* **484**, 10 (2000).
12. M. L. Ambriola *et al.*, *Nucl. Phys. (Proc. Suppl.) B* **78**, 32 (1999).
13. D. Bergström *et al.*, *Nucl. Instr. and Meth. A* **463**, (2001) 161.
14. M. Bocciolini *et al.*, *Nucl. Instr. and Meth. A* **370**, (1996) 403.
15. M. Boezio *et al.*, *Astrophys. J.* **532**, 653 (2000).
16. A. S. Beach *et al.* astro-ph/0111094.
17. R. L. Golden *et al.*, *Astrophys. J.* **457**, L103 (1994).
18. S. W. Barwick *et al.*, *Astrophys. J.* **498**, 779 (1998).
19. M. Boezio¹, Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden (1998).
20. J. D. Sullivan, *Nucl. Instr. and Meth.* **95**, (1971) 5.
21. A. S. Stephens, in Proc. 17th Int. Cosmic-Ray Conf., **4**, 282 (Commissariat a l'Energie Atomique, Gif-sur-Yvette, Essonne, France, 1981)

SUPERNOVA REMNANTS AS HADRONIC COSMIC RAY ACCELERATORS?

Yousaf M. Butt¹, Diego F. Torres^{2,3}, Jorge A. Combi³, Thomas Dame¹,
Gustavo E. Romero³

¹*Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge,
MA, USA.*

²*Department of Physics, Princeton University, Princeton, NJ, USA*

³*Instituto Argentino de Radioastronomía, C.C. 5, 1894, Villa Elisa, Buenos
Aires, ARGENTINA*

ABSTRACT

The non-thermal supernova remnant (SNR) RX J1713.7-3946 has recently been shown to be a site of cosmic ray (CR) electron acceleration to TeV energies (Muraishi et al., 2000). Here we present evidence that this remnant is also accelerating cosmic ray *nuclei*. Such energetic nuclei can interact with ambient interstellar gas to produce high energy gamma-rays via the decay of neutral pions. We associate the unidentified *EGRET* GeV gamma-ray source, 3EG J1714-3857, with a very massive ($\sim 3 \cdot 10^5 M_{\odot}$) and dense (~ 500 nucleons cm^{-3}) molecular cloud interacting with SNR RX J1713.7-3946. Direct evidence for such interaction is provided by observations of the lowest two rotational transitions of CO molecules in the cloud; as in other clear cases of interaction, the $\text{CO}(J=2 \rightarrow 1)/\text{CO}(J=1 \rightarrow 0)$ ratio is significantly enhanced. Since the cloud is of low radio and X-ray brightness, electrons cannot be responsible for the bulk of the GeV emission there. A picture thus emerges where both electrons and nuclei are being accelerated by the SNR: whereas the relativistic electrons dominate the *local* non-thermal radio, X-ray and TeV emission, the shock accelerated CR protons and ions (hadrons) are exposed through their interactions in the *adjacent* massive cloud, leading to the observed GeV emission via the gamma-decay of neutral pions. Such a scenario had been anticipated by Aharonian, Drury and Völk (1994).

1. INTRODUCTION

The question over the origin of cosmic rays (CR) has persisted ever since they were first detected in 1912 by Victor Hess using a balloon-borne electroscope device (Hess, 1912). Despite the fact that there is now broad consensus that the shocks in the expanding blast waves of supernova remnants (SNRs) accelerate the bulk of the CR ions up to energies of ~ 300 TeV per nucleon, and possibly to even $\sim 10^5$ TeV per nucleon (Bell & Lucek, 2001), direct evidence of this scenario has so far eluded observers. The best way to identify such energetic CR accelerators is to look for the associated high-energy (>100 MeV) gamma-rays produced at the sources; however, since CR electrons and nuclei can both generate gamma-rays at the acceleration sites, it has not yet been possible to unambiguously associate the detected gamma-radiation with sources of CR *nuclei*, specifically. In fact, only very recently have ground-based imaging air _erenkov gamma-ray telescopes provided direct evidence for the presence of highly relativistic, TeV energy *electrons* at the shocks of three shell-type SNR's: SN1006 (Tanimori et al., 1998), Cassiopeia A (Aharonian et al., 2001) and RX J1713.7-3946 (also known as G347.3-0.5; Muraishi et al., 2000), the remnant considered here. Even though previous reports (eg. Sturmer et al., 1996; Esposito et al., 1996; Gaisser et al., 1998; Aharonian & Atoyan, 1999; Combi et al., 1998, 2001) have suggested that the gamma-ray emission seen in the directions of some SNRs could be due to hadronic interactions, it has not been possible to rule out energetic electrons or the nearby pulsars as the dominant source of the detected radiation (eg. Brazier et al., 1996; De Jager, & Mastichiadis, 1997; Gaisser et al., 1998).

2. 3EG J 1714-3857 AND SNR RX J1713.7-3946

Prompted by the close association of the GeV *EGRET* source 3EG J1714-3857 (Hartman et al., 1999) with the TeV gamma-ray emitting SNR RX J1713.7-3946, we have investigated the possibility that this remnant is also accelerating nuclei, *in addition to the known CR electron acceleration taking place there* (Muraishi et al., 2000). As Figure 1 shows, there are two massive and dense molecular clouds lying adjacent to the SNR; one of which, Cloud A, immediately abuts the blast wave region of the SNR (Slane et al., 1999). This cloud is also partially within the inner 50% confidence location contour of 3EG J1714-3857 (Hartman et al., 1999). Both clouds have a mean Local Standard of Rest (LSR) velocity of -94 km/sec and an inferred kinematic distance of 6.3 ± 0.4 kpc. The mass of Cloud A, is determined from CO observations (Bronfman et al., 1989) to be $(3 \pm 0.3) \times 10^5 M_{\odot}$, and its mean density ~ 500 nucleons cm^{-3} . Cloud B has a mass of $(2.8 \pm 0.3) \times 10^5 M_{\odot}$, and mean density of ~ 660 nucleons cm^{-3} .

As the blast wave of RX J1713.7-3946 overtakes Cloud A the shock-accelerated protons and ions collide with the resident nuclei and produce neutral pions which then promptly gamma-decay ($\pi^0 \rightarrow \gamma\gamma$), illuminating the cloud at GeV energies; such a scenario has been anticipated by several authors (eg. Montmerle, 1979; Aharonian, Drury & Volk, 1994; Dorfi, 1991, 2000). Indeed, a recent detailed analysis of the broadband *electronic* emissions of RX J1713.7-3946 by Ellison et al. (2001) directly supports our findings by suggesting that 25-50% of the forward shock kinetic energy is likely being taken up in accelerating ions to relativistic energies of up to ~ 70 TeV/nucleon in this remnant. Strong evidence that the shock front of RX J1713.7-3946 has overtaken, and is interacting with Cloud A, is provided by the enhanced intensity ratio of the two lowest rotational transitions of the CO molecules in the cloud, as already noted by Slane et al. (1999). This ratio, $R = \{CO(J=2 \rightarrow 1)/CO(J=1 \rightarrow 0)\}$, is typically ~ 0.7 in the Galactic plane (Sakamoto et al., 1995), but is known to be enhanced in shocked molecular gas interacting with SNRs (eg. Seta et al., 1998). We examined this ratio, R , at all LSR velocities over a $2^\circ \times 1^\circ$ region centered roughly on Cloud A, using CO($J=1 \rightarrow 0$) data from Bronfman *et al.* (1989), and unpublished CO($J=2 \rightarrow 1$) data from the University of Tokyo 0.6m telescope at La Silla, Chile, kindly provided by T. Handa and T. Hasegawa. As Figure 2 shows, of all the 781 ratios measured, 2 of the 3 highest (top 0.5% percentile), with $R \sim 2.4 \pm 0.9$, were found on or very close to Cloud A in both position and velocity. In contrast, no enhanced CO($J=2 \rightarrow 1$)/CO($J=1 \rightarrow 0$) ratio was observed in the vicinity of Cloud B.

We have calculated the expected gamma-ray luminosity for the proposed scenario using the following information from Slane et al. (1999): supernova explosion energy = $E_{SN} = (1.7 - 2.2) \times 10^{51}$ ergs; distance to the SNR = 6.3 ± 0.4 kpc; unshocked ambient density, $n_o = 0.01 - 0.3 \text{ cm}^{-3}$; together with the cloud data extracted above. Using the Sedov solution we calculate the age of the SNR to be in the range $(2.5 - 13.4) \times 10^4$ yrs. The total gamma-ray luminosity is divided between that from the hadronic interactions intrinsic to the SNR, and that due to the enhanced probability of hadronic interactions in the high target density medium of Cloud A:

$$F_{tot}(E > 100 \text{ MeV}) = F_{snr}(E > 100 \text{ MeV}) + F_{cloud A}(E > 100 \text{ MeV})$$

We may evaluate the first term as (Drury et al., 1994):

$$F_{snr}(E > 100 \text{ MeV}) \sim 4.4 \times 10^{-7} \theta E_{51} D_{kpc}^2 n_o$$

where θ is the fraction of the total supernova energy converted to cosmic ray energy (eg. Morfill et al., 1984); E_{51} is the supernova explosion energy in units of 10^{51} erg; and D_{kpc} is the distance in kpc.

If the start of the Sedov phase is taken at ~ 2400 years, then substituting the numerical values for the various quantities yields the intrinsic GeV luminosity of the SNR to be in the range:

$$F_{snr}(E>100\text{MeV}) = (0.1-3) \times 10^{-9} \text{ photons cm}^{-2} \text{ sec}^{-1}$$

The second term represents the contribution to the GeV flux from the SNR amplified CR bombardment of Cloud A and is given by (Aharonian & Atoyan, 1996):

$$F_{cloud A}(E>100\text{MeV}) = 2.2 \times 10^{-7} M_5 D_{kpc}^2 k_s \text{ photons cm}^{-2} \text{ sec}^{-1}$$

where M_5 is the mass in units of $10^5 M_o$ and k_s is the cosmic ray enhancement factor, ie. the ratio of the CR energy density in the vicinity of the SNR to that measured near the Sun. [We adopt a gamma-ray emissivity, $q(E>100\text{MeV})=2.2 \times 10^{-25}$ photons (H-atom) $^{-1} \text{ sec}^{-1}$, (Dermer, 1986)]. Using Morfill et al. (1984), we find $24 < k_s < 36$ for the SNR at the current epoch. Thus, the GeV luminosity of Cloud A is, $F_{cloud A}(E>100\text{MeV}) = (3.1 - 7.6) \times 10^{-7}$ photons $\text{cm}^{-2} \text{ sec}^{-1}$, and dominates, by more than two orders of magnitude, the GeV flux produced by the SNR itself. Since the intrinsic SNR contribution can thus be neglected, the total expected hadronically generated gamma-ray luminosity is simply the same as that from Cloud A:

$$F_{tot}(E>100\text{MeV}) = (3.1 - 7.6) \times 10^{-7} \text{ photons cm}^{-2} \text{ sec}^{-1}$$

That this predicted flux is fully consistent with the measured value, $(4.36 \pm 0.65) \times 10^{-7}$ photons $\text{cm}^{-2} \text{ s}^{-1}$ (Hartman et al., 1999), further supports a nucleonic source of the detected gamma rays. Also, the fact that the *EGRET* source is coincident with the molecular cloud, and not the SNR itself, is fully in agreement with our calculation of the expected fluxes from these two sources. In addition, the spectral index of the GeV source, $\Gamma = -2.3 \pm 0.2$ (Hartman et al., 1999), is in tune with that expected from the hadronic interactions of a *source* CR population (Fields et al., 2001).

To be certain of a nucleonic source of the detected GeV flux from Cloud A, however, it is crucial to eliminate the alternative, electromagnetic origin of the gamma-rays. We show that the *non*-detection of Cloud A in the radio band (Slane et al., 1999; Ellison et al., 2001) rules out the possibility that electrons are contributing significantly to the GeV luminosity of the cloud. At the high particle densities of Cloud A, the contribution of the electron IC process to the GeV luminosity can be neglected in comparison to the electron bremsstrahlung process (eg. De Jager & Mastichiadis, 1997). However, the electron flux needed to explain the intensity of the measured GeV emission via electron bremsstrahlung in the cloud material will produce an enhanced radio emission by the

synchrotron mechanism which far exceeds the measured values. The expected ratio of gamma-ray (>100 MeV) electron bremsstrahlung flux to the radio synchrotron flux may be expressed:

$$R = \frac{F(E > 100 \text{ MeV})}{F(\nu)_{\text{Jy}}} = \frac{4.3 \times 10^{-21}}{c(p)} n_{\text{cm}^{-3}} B_{\mu\text{G}}^{-(1+p)/2} \nu_{\text{Hz}}^{(p-1)/2} \text{Jy}^{-1} \text{cm}^{-2} \text{s}^{-1},$$

where,

$$c(p) = 10^{-5(1+p)} (3.2 \times 10^{15})^{(p-1)/2} (p-1)^2 a(p),$$

$a(p)$ is given in Longair (1994), and $p=2.3$ is the spectral index of the electron population, $N_e(E) \sim E^{-2.3}$

Then, using the physical parameters of Cloud A, together with an assumed magnetic field of 4×10^{-5} G [a conservative estimate considering the density of the medium (Crutcher et al., 1987)], we can calculate that the predicted radio luminosity at 843 MHz – *under the assumption of an electronic origin of the GeV flux* – would be ~ 85 Jy. Since this flux is about 20 times larger than the *upper limit* derived from the non-detection of Cloud A at this frequency (Slane et al., 1999), we conclude that no significant part of Cloud A’s GeV radiation could be due to electronic processes [Fig 3]. Furthermore, were the GeV flux of 3EG J1714-3857 of electronic origin, Cloud A would outshine even the radio-brightest NW rim of the remnant, which is found to be emitting at only 4 ± 1 Jy at 1.36 GHz (Ellison et al., 2001). Thus, clearly, electrons are not responsible for the bulk of the measured GeV flux of 3EG J1714-3857. There is, of course, a relativistic bremsstrahlung contribution from *secondary* electrons and positrons produced by the decay of *charged* pions ($\pi^\pm \rightarrow \mu^\pm \rightarrow e^\pm$) which are also generated in the hadronic interactions, but this gamma-ray intensity is expected to be more than an order of magnitude lower than the $\pi^0 \rightarrow \gamma\gamma$ flux above 100 MeV (Berezinskii et al., 1990).

3. OTHER SOURCES IN THE FIELD

There are also two other SNRs projected within 3EG J1714-3857’s 95% contours, CTB37A&B. However, since these SNRs are more distant (11.3 kpc), and because their *maximum possible* interacting cloud mass (Reynoso & Mangum, 2000) is measured to be an order of magnitude less than Cloud A’s, their contribution to the GeV luminosity is less than 2×10^{-8} photons $\text{cm}^{-2} \text{sec}^{-1}$, or less than 5% of the measured 3EG J1714-3857 flux.

The two pulsars within the 95% confidence location contours of this *EGRET* source, PSR J1715-3903 and J1713-3844 (Manchester et al., 2001), can also be eliminated as the source of the bulk of

the measured GeV flux. PSR J1713-3844 at $(l,b)=(348.10,+0.21)$ is a long-period pulsar ($P=1.60011$ sec) whose spin-down luminosity is two orders of magnitude below that needed to account for 3EG J1714-3857. Although spinning faster, PSR J1715-3903 at $(l,b) = (348.10, -0.32)$, is still not energetic enough to be responsible for 3EG J1714-3857: it has a period $P=0.27848$ sec; the dispersion measure indicates a distance of $d=4.8$ kpc and the observed period derivative of 37.688×10^{-15} implies a spin-down luminosity of $\dot{E} \sim 7 \times 10^{34}$ erg s^{-1} , for a standard neutron star moment of inertia $I=10^{45}$ g cm^2 . Thus, $\dot{E}/d^2 \sim 3 \times 10^{33}$ erg s^{-1} kpc^{-2} which is more than an order of magnitude below the lowest value among the confirmed gamma-ray pulsars (Kaspi et al., 2000). [We do not consider PSR B1055-52 since there is an open controversy regarding the distance to this pulsar – see Combi et al. (1997), Romero (1998) and Mc Laughlin & Cordes (2000) for discussions.] We thus conclude that the pulsar J1715-39 is not responsible for the bulk of the GeV emission of 3EG J1713.7-3946. This conclusion is supported by the lack of any X-ray counterpart of PSR J1715-39 in any archival X-ray database, including the *ROSAT* all-sky survey.

Lastly, the *EGRET* source 3EG J1713.7-3946 is not coincident with any other candidate gamma-ray sources such as OB associations, Wolf-Rayet or Of stars (Romero et al., 1999; Torres et al., 2001). The analyses of both Tompkins (1999) and Torres et al. (2001) also shows this source to be non-variable, as should be the case for an interacting SNR.

4. CONCLUSIONS

We have argued that the unidentified *EGRET* source 3EG J1714-3857 (Hartman et al., 1999) results predominantly from the gamma-rays produced by nuclei accelerated by SNR RX J1713.7-3946 interacting with those resident in the dense and massive molecular cloud immediately abutting the remnant. A recent analysis of the *electronic* emissions of this remnant by Ellison et al. (2001) directly supports our proposal by suggesting that 25-50% of the forward shock kinetic energy is likely being taken up in accelerating ions to relativistic energies of up to ~ 70 TeV/nucleon.

However, it should be noted that a number of theoretical assumptions are built into the simple models of the literature cited in our analysis of the hadronic gamma-ray production (Morfill et al., 1984; Aharonian et al., 1994; Drury et al., 1994; Aharonian & Atoyan, 1996). For instance, the diffusion and confinement of the protons in the dense and magnetized media of molecular clouds is a complicated problem whose detailed analysis is beyond the scope of this letter (see, eg. Zweibel & Shull, 1982; Berezhinskii et al., 1990; Dogiel & Sharov, 1990; Chandran, 2000). The large angular size of the *EGRET* error box [Fig. 1] also leaves open the possibility that some other, as yet

unidentified, source could also be contributing significant gamma-ray flux. Observations of this region with the upcoming higher sensitivity and spatial resolution satellite-based GeV telescopes, such as *AGILE* and *GLAST*, will thus be very important. We also propose that Cloud A of RX J1713.7-3946 be a high priority target for *CANGAROO-III* and *HESS*, the forthcoming high-sensitivity ground-based TeV _erenkov telescope arrays in the southern hemisphere. Such observations of very high energy photons from Cloud A could directly probe the maximum proton energy, E_{p-max} , accelerated by RX J1713.7-3946, since the hadronic gamma-ray spectrum begins steepening at $E_\gamma \sim 0.1 E_{p-max}$ and is cut-off at $E_\gamma \sim E_{p-max}$ (eg. Naito & Takahara, 1994).

In conclusion, the facts that TeV energy cosmic ray electrons are accelerated in SNR RX J1713.7-3946 (Muraishi et al., 2000); that the abutting cloud material is inordinately excited; that the cloud region is of low radio and X-ray brightness; that the GeV luminosity is non-variable and in quantitative agreement with that expected from π^0 gamma-decays; that the spectral index is as expected for an hadronic CR source population; and, lastly, that there are no other known candidate sources within the 95% location contours of 3EG J1714-3857 capable of explaining the GeV flux, all suggest that this *EGRET* source is the gamma-ray signature of accelerated nuclei from SNR RX J1713.7-3946 interacting with those of the neighboring dense and massive molecular cloud.

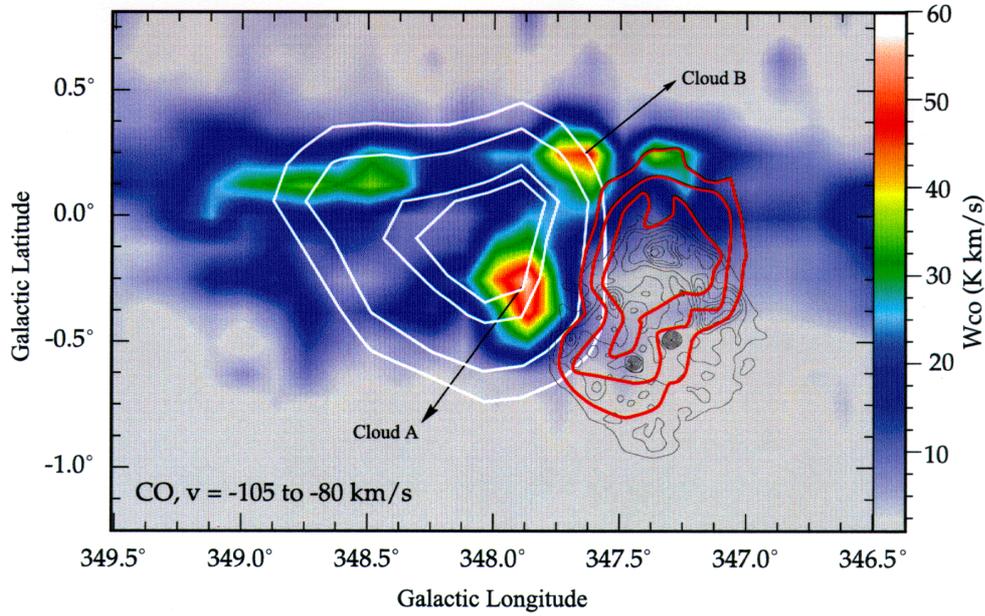


Figure 1: An overlay map in Galactic coordinates showing SNR RX J1713.7-3946 (G347.3-0.5) in grey (ROSAT PSPC X-ray) contours from Slane et al. (1999). Red depicts the TeV significance contours from Muraishi et al. (2000). In white are the location probability contours (successively, 50%, 68%, 95% and 99%) of the GeV EGRET source 3EG J1714-3857 from Hartman et al. (1999). The color-scale indicates the intensity of CO(J=1→0) emission, and consequently the column density of the ambient molecular cloud, in the LSR velocity interval $v_{\text{LSR}} = -105$ to -80 km/sec associated with the SNR, corresponding to a kinematic distance of 6.3 ± 0.4 kpc. The elongated CO emission feature near (l,b)~(348.5,+0.2) derives from the large velocity wings of a much more distant (~11.3 kpc) and unrelated cloud centered at $V_{\text{LSR}} = -68$ km/sec.

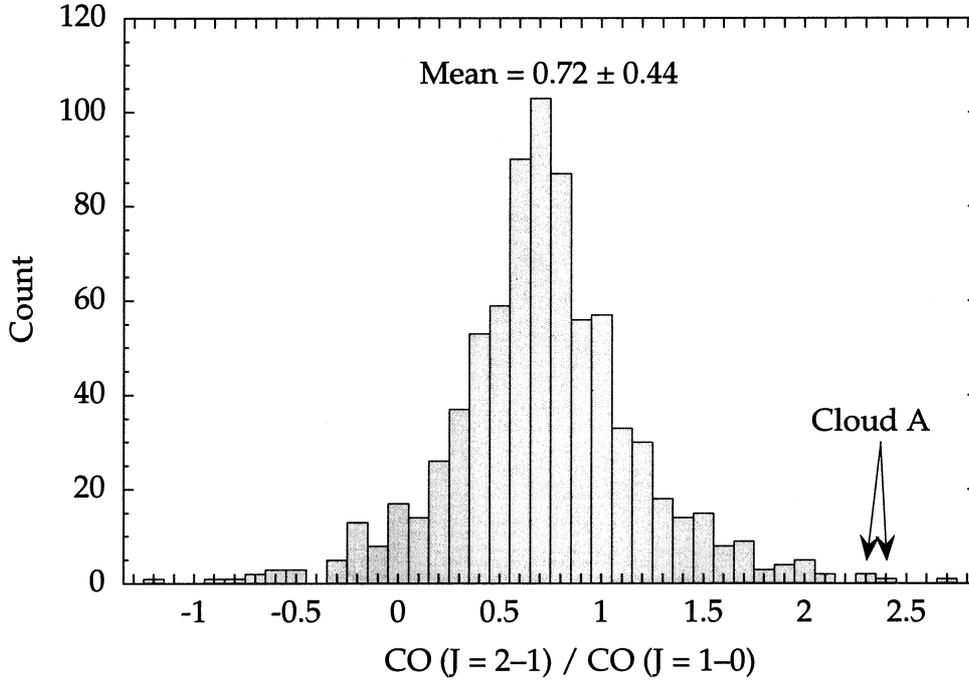


Figure 2: The distribution of all 781 line intensity ratios, $R = \{CO(J=2 \rightarrow 1)/CO(J=1 \rightarrow 0)\}$, measured every 15_ in the region from $l=346.5 \rightarrow 348.5$; $b = -0.5 \rightarrow +0.5$, and averaged over 5km/sec bins of velocity between $v_{lsr} = -150$ km/sec \rightarrow +50 km/sec. Pixels in which the $CO(J=1 \rightarrow 0)$ intensity is less than 2.5 times the instrumental noise are excluded. The bins labeled “Cloud A” contain 3 pixels, 2 of which are consistent with the position **and** velocity of Cloud A: $(l,b,v)=(348.0,-0.25,-85$ km/s), and $(l,b,v)=(348.25,-0.5,-90$ km/s). All other pixels with high R values ($R > 1.8$) lie well outside the 95% confidence location contour of 3EG J1714-3857. The mean of the distribution, ~ 0.72 , agrees with the average unexcited value in the Galactic plane (Sakamoto et al., 1995). The dispersion about the mean of ~ 0.44 , results both from the intrinsic scatter in R, as well as from instrumental noise, mainly in the $CO(J=2 \rightarrow 1)$ data. That the latter source dominates is evidenced by some unphysical negative R values which are caused by background subtraction in pixels with very low $CO(J=2 \rightarrow 1)$ intensity.

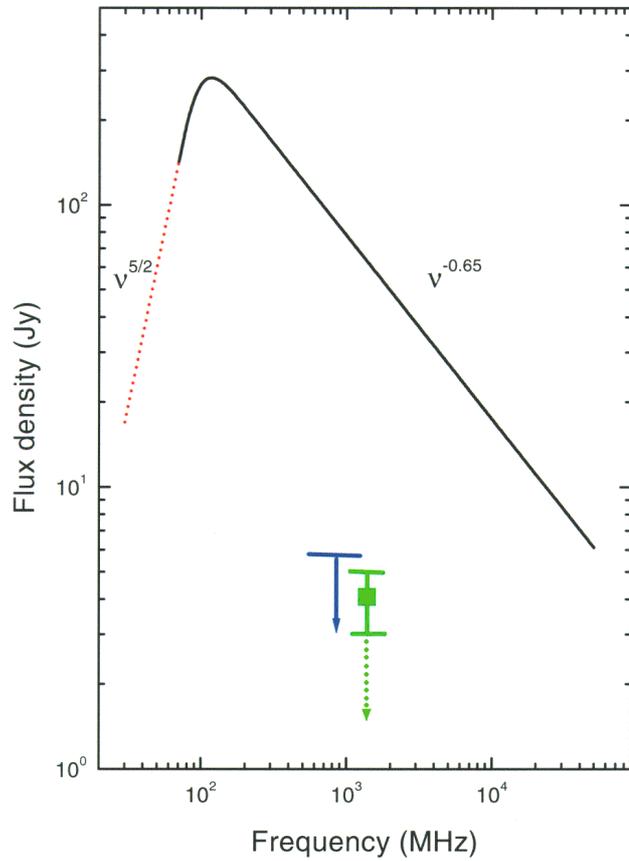


Figure 3: The radio synchrotron spectrum that would be expected from Cloud A (located towards the N.E. of the remnant) if the GeV flux were due to either primary or secondary electron/positron bremsstrahlung (see text for details). Since this spectrum violates the upper limit (blue) derived from the non-detection of Cloud A in the radio band by a factor of ~ 20 at 843 MHz (Slane et al., 1999), we rule out a predominantly leptonic origin of the GeV luminosity. Furthermore, were the GeV flux of 3EG J1714-3857 of electronic origin, Cloud A would outshine even the radio-brightest N.W. rim of the remnant which is found to be emitting at only 4 ± 1 Jy at 1.36 GHz (Ellison et al., 2001), as shown by the green datapoint. This latter datapoint from the N.W. rim of the remnant thus constitutes another effective upper limit, shown by the green dots, for the radio emission of Cloud A, at a frequency of 1.36 GHz. (An assumed low frequency turnover at ~ 100 MHz is shown by the red dotted line.)

ACKNOWLEDGMENTS

We thank T. Handa and T. Hasegawa for kindly providing the CO($J=2\rightarrow 1$) data, obtained at the University of Tokyo 0.6m telescope at La Silla, Chile. We are grateful to Patrick Slane for alerting us to RX J1713.7-3946 and to Don Ellison, Sera Markoff and David Thompson for useful discussions and information. The *EGRET* and HEASARC archives at the Goddard Space Flight Center; the on-line Australia Telescope National Facility (ATNF) pulsar archive; as well as the *ROSAT* all-sky survey from the Max-Planck Institute were all invaluable to this study. The Relativistic Astrophysics Group at IAR (DFT, JAC & GER) is supported by *CONICET*, *ANPCT*, and Fundación Antorchas. YMB is supported by the High Energy Astrophysics Division and the *CHANDRA* project at the *CfA*; NASA contract NAS8-39073.

REFERENCES

- Aharonian F. A., Drury, L. O'C., Völk, H. J. 1994, *A&A*, **285**, 645
Aharonian, F. A. & Atoyan, A. M. 1996, *A&A*, **309**, 917
Aharonian, F. A., & Atoyan, A. M. 1999, *A&A*, **351**, 330
Aharonian, F. *et al.* 2001, *A&A*, **370**, 112
Berezinskii, V. S., Bulanov, S. V., Dogiel, V. A., Ptuskin, V. S., *Astrophysics of Cosmic Rays*, Amsterdam: North-Holland (1990), edited by V. L. Ginzburg.
Brazier, K. T. S. *et al.* 1996, *MNRAS*, **281**, 1033
Bronfman, L., *et al.* 1989, *ApJS*, **71**, 481
Chandran, B. D. G. 2000, *ApJ*, **529**, 513
Combi, J. A., Romero, G. E., Azcarate, I. N. 1997, *Ap&SS*, **250**, 1
Combi, J. A., Romero, G. E., Benaglia, P. 1998, *A&A*, **333**, L91
Combi, J. A. *et al.* 2001, *A&A*, **366**, 1047
Crutcher, R. M., Troland, T. H., Kazes, I. 1987, *A&A*, **181**, 119
De Jager, O. C. & Mastichiadis, A. 1997, *ApJ*, **482**, 874
Dermer, C. D. 1986, *A&A*, **157**, 223
Dogiel, V. A. and Sharov, G. S. 1990, *A&A*, **229**, 259
Dorfi, E. A. 2000, *Ap&SS*, **272**, 227
Dorfi, E. A. 1991, *A&A*, **251**, 597
Drury, L. O'C., Aharonian, F., Völk, H. J. 1994, *A&A*, **287**, 959
Ellison, D., *et al.* 2001, "Broad-band Observations and Modelling of the Shell-Type Supernova Remnant G347.3-0.5", *submitted to ApJ*, **astro-ph/0106257**
Esposito, J. A., *et al.* 1996, *ApJ*, **461**, 820
Fields, B. D., *et al.* 2001, *A&A*, **370**, 623
Fierro, J. M., *et al.* 1993, *ApJ*, **413**, L27
Gaisser, T. K., Protheroe, R. J., Stanev, T. 1998 *ApJ*, **492**, 219
Hartman, R. C., *et al.* 1999, *ApJS*, **123**, 79

- Hess, V. F. 1912, *Physik Z.*, **13**, 1084
- Kaspi, V., et al. 2000, *ApJ*, **528**, 445
- Longair, M. S., *High Energy Astrophysics. Vol.2: Stars, the Galaxy and the Interstellar Medium*, Cambridge University Press, (1994), 2nd ed.
- Manchester, R. N. *et al.*, *MNRAS*, in press (2001), **astro-ph/0106522**, see on-line catalog at <http://wwwatnf.atnf.csiro.au/research/pulsar/pmsurv/>
- Mc Laughlin, M. A., and Cordes, J. M. 2000, *ApJ*, **538**, 818
- Montmerle, T. 1979, *ApJ*, **231**, 95
- Morfill, G. E., Forman, M., & Bignami, G. 1984, *ApJ*, **284**, 856-868
- Muraishi, H., *et al.* 2000, *A&A*, **354**, L57
- Naito, T. & Takahara, F. 1994, *J Phys. G*, **20**, No. 3, 477
- Reynoso, E., Mangum, J. G. 2000, *ApJ*, **545**, 874
- Romero, G. E. 1998, *Rev. Mex. Astron. Astrofis.*, **34**, 29
- Romero, G. E., Benaglia, P., Torres, D. F. 1999, *A&A* **348**, 868
- Sakamoto, S., et al. 1995, *ApJS*, **100**, 125
- Seta, M., *et al.* 1998, *ApJ*, **505**, 286
- Slane, P., *et al.* 1999, *ApJ*, **525**, 357
- Sturmer, S. J., Dermer, C. D., & Mattox, J. R. 1996, *A&AS*, **120**, 445
- Tanimori, T., *et al.* 1998, *ApJ*, **497**, L25
- Tompkins, W. 1999, *Applications of Likelihood Analysis in Gamma-Ray Astrophysics*, Ph.D. Thesis, Stanford University
- Torres, D. F., et al. 2001, Positional Correlation Between Low-Latitude Gamma-Ray Sources and SNRs, *to be published in the proceedings of the workshop: The Nature of the Unidentified Galactic Gamma-Ray Sources*, INAOE, Mexico, October 2000. (A. Carraminana, O. Reimer and D. Thompson Eds.), **astro-ph/0012160**
- Torres, D. F., et al. 2001, *A&A*, **370**, 468
- Zweibel, E. & Shull, J. M. 1982, *ApJ*, **259**, 859

RADIATION IN SPACE: CAUSES, MEASURES AND COUNTERMEASURES

M. Casolino
INFN and University of Tor Vergata, Roma

ABSTRACT

The solar - terrestrial environment is a time dependent system in which different phenomena at different energies play very interrelated roles. The Sun is the dominant source of energy not only of electromagnetic radiation but also of charged cosmic rays. The energy of these particles goes from the low energy solar wind to the high energy Solar Energetic Particle Events. These particles interact both with Galactic Cosmic Rays and with the Earth's magnetosphere, producing an environment being studied from different points of view. In this work we briefly cover the principal causes for radiation in interplanetary space and within Earth's magnetosphere, analyzing the amount and effects that cosmic rays have on human activities on Earth and in space.

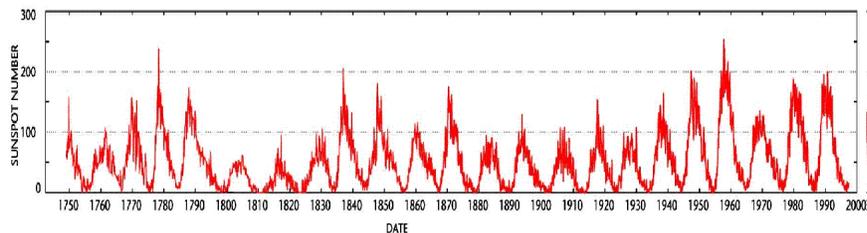


Figure 1: *Sunspot number as a function of time.*

1 The Solar Terrestrial Environment

The sun is the dominant source of energy in the solar system: it emits radiation in the form of electromagnetic waves from the lowest radio frequencies to $\approx MeV$ γ rays; is the source of low energy (KeV - GeV) cosmic rays¹. The energy comes from nuclear fusion processes occurring at the *Core* ($R < 0.25R_{sun} = 696000km$) and flows through the *Radiative zone* ($0.25R_{sun} < R < 0.7R_{sun}$) to the *Convective zone*, which is surmounted by the *Photosphere*, followed by the *Chromosphere* and the expanding *Corona*. If the core and the radiative zones can be described by relatively few parameters, the convection zone and the upper layers are much more complex to model and describe due to the amount of different processes that take place. The field is so vast that the reader is referred to ^{1, 2, 4}) and references therein. One of the most striking features of the sun is the presence of many time-dependent phenomena with extremely varying time scales according to the processes involved. We mention the *solar oscillations* ⁶), periodic acoustic resonance waves (main modes have $\tau = 5, 20 - 60$ and $160 min$) which have provided important information on the interior of the Sun. On a much larger scale, both temporal and spatial, we observe the *solar cycle*, with a main periodicity of 11 years. The detection of solar cycle was related to the first observations of sunspots, regions often larger than Earth where the magnetic field, being much more intense ($B \approx 0.3T$) than the average solar magnetic field ($B \approx 10^{-4}T$) inhibits plasma flow and thus

¹The highest cosmic ray flux on Earth is due to neutrinos produced in nuclear fusion processes at the core of the Sun: on Earth the flux from the pp reaction is $0.4 \times 10^{11} \nu/s$ ¹).

lowers its temperature, resulting in a lower e.m. emission. The first systematic observations of sunspots started in the 17th century but a clear evidence of an 11-year cycle (Figure 1) of the observed sunspot number (ssn) was delayed until the beginning of the 18th century due to the fact that the years between 1630 and 1710 were marked by an extremely limited amount of sunspots². Now the 11-year solar cycle is currently explained in terms of a periodic reversal of the solar magnetic field from a solar minimum (when ssn is low and the magnetic field is dipolar), passing through a solar maximum (when ssn is high and the magnetic field is scattered and irregular) to the next solar minimum (when the magnetic field is again dipolar but with inverted polarity); a complete cycle thus takes 22 years to complete. The solar cycle is currently explained in terms of a *dynamo effect*, powered by differential solar rotation: the sun is not a rigid body and thus has a latitude (θ) dependent rotation speed:

$$\omega = \frac{2\pi}{26} \left(1 - \frac{1}{8} \sin^2 \theta - \frac{1}{6} \right) \text{ rad/day} \quad (1)$$

The ionized plasma of the sun's surface is linked to the magnetic field lines; starting from a dipolar structure at solar minimum with field lines parallel to the sun meridians, the different rotation speed - faster at the equator, causes the lines to distort assuming with time an almost horizontal shape. The lines also become twisted among themselves, increasing the local magnetic field intensity; when they surface due to buoyancy processes they produce sunspots. At solar maximum the field is no longer dipolar and is gradually disrupted by convection processes to reform again in an dipolar shape with its polarity inverted.

1.1 The Solar Wind

Of particular relevance for the study of the solar - terrestrial environment is the investigation of the solar wind. The high temperature ($10^5 K$) solar corona is not gravitationally bound due to its high kinetic and therefore expands in the interplanetary medium. This expansion gives rise to the solar wind, a totally ionized gas composed mostly by protons and electrons which blows in the

²The so-called Maunder minimum: This period, as the previous (Sporer: 1450-1540) and the following (Dalton: 1795-1520) minima resulted in a series of climatical changes on Earth and coincide with the colder periods of the little Ice Age (1450-1820).

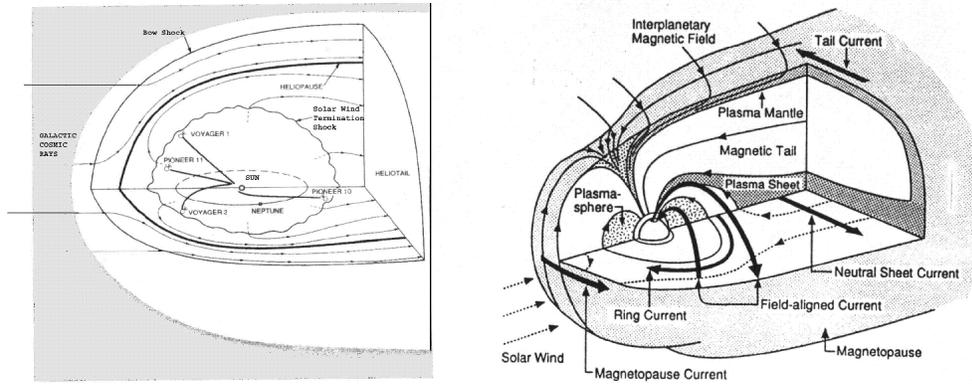


Figure 2: *Left: Sketch of the heliosphere. The relative motion of the solar system within the galaxy creates a bow shock which partially deflects incoming particles. The heliopause, where the interstellar and interplanetary magnetic field have the same intensity is located between the bow shock and the solar wind termination shock, where solar wind particle speed becomes subsonic (Adapted from 11). Right: A similar structure forms when the solar wind interacts with the Earth's magnetosphere, giving rise to a magnetosphere bow show. The magnetopause, where interplanetary and terrestrial magnetic fields are equal, is the source of several particle currents 2).*

whole solar system. Through the phenomenon of *flux freezing* the interplanetary magnetic field is linked to the solar wind particles 2). Since the kinetic energy density $\epsilon_{kinetic} = \rho v^2/2 = 0.6 \text{ erg cm}^{-2} \text{ s}^{-1}$ of the plasma is higher than the magnetic field energy density $\epsilon_{mag.field} = B^2/(2\mu_0) = 0.01 \text{ erg cm}^{-2} \text{ s}^{-1}$ the radially expanding solar wind carries away the magnetic flux; however the 26 days solar rotation causes a distortion of the field lines, which assume an archimedean spiral configuration. This distortion is such that at 1AU the magnetic lines have an angle of about 45° in respect to the sun-Earth line 4). The solar wind continues its expansion until it interacts with the galactic magnetic field: indeed the boundary of our solar system can be considered divided in three layers. Toward the outside we have (Figure 2-left:

1. *the solar wind termination shock*, where the gradual reduction of the speed of the wind comes to an abrupt end due to the passage from su-

personic to subsonic speeds due to the interstellar background pressure ($\approx 10^{-13}$ Pa))

2. *the heliopause*, where the values of the interplanetary and interstellar magnetic fields are equal ($2.5 \times 10^{-10}T$)
3. *the bow shock*, the region of subsonic slow down of the *interstellar* medium.

Although these regions have not yet been observed, (the Voyager 2 is the probe closest to the end of the solar system but hasn't come across the termination shock yet), there is ample evidence (for instance from cosmic ray modulation) for its existence and its location between 100 and 150 AU.

The structure of the end of the solar system, where the interplanetary and interstellar media interact, is reproduced in a much smaller scale close to the Earth, with the geomagnetic and the interplanetary fields.

1.2 The Earth's magnetosphere

Another element in the solar terrestrial environment is the geomagnetic field and its interaction with the interplanetary magnetic field carried by the solar wind. Our magnetic field can be considered - at a first approximation - as an inverted dipole of $30.4\mu T R_{earth}$ tilted 11° in respect to the rotation axis and offset in respect to the center of the Earth. More correctly the geomagnetic field may be expressed by a Legendre (spherical harmonic) series with internal ($\propto \frac{1}{R^n}$) and external ($\propto R^n$) terms²). These terms are not constant: intensity and direction of the terrestrial magnetic field varies over the years³ In addition to secular variations there are also transient phenomena related to solar activity and the interaction with the charged particles of the solar wind. As shown in Figure 2-right, the dipolar structure of the magnetic field is distorted by the interaction with solar wind particles: in analogy to the heliopause, there is the formation of a bow shock at a distance given by:

$$R_{bow} = \frac{25R_{Earth}}{1 + 0.8 \cos\theta} \quad \theta = 0 \rightarrow \textit{sunward looking} \quad (2)$$

on the night side this equation is not valid: the flow of the solar wind elongates the geomagnetic field lines up to more than $200 R_{earth}$ (the distance

³Topographical maps, for instance, show the local direction and first derivative of the magnetic North to allow for this correction.

between the Moon and Earth is about $60 R_{earth}$) producing the geomagnetic tail. Inside the bow shock we find the magnetopause where the magnetic field intensities of the interplanetary medium and the Earth are equal ($\approx 5 \times 10^{-9} T$). Inside the magnetosphere there is the obvious absence of the analogue of the solar wind termination shock. Our magnetosphere is the only region where is possible to study with *in situ* measurement collisionless shocks and the phenomena related to the interaction between solar wind and the geomagnetic field. Knowledge gathered directly in this regions can be applied not only to an understanding of the heliopause but also to modelling all effects where magnetohydrodynamics plays an important role: from solar energetic particle acceleration to supernova shocks to interaction between particles and neutron star magnetic fields.

The geomagnetic field deflects also galactic cosmic rays: the *geomagnetic cutoff* G represents the impulse below which an orthogonally incident particles of charge z cannot reach the Earth:

$$G = 14.9 z \cos^4 \lambda (Gev/c) \quad (3)$$

where λ represents the geomagnetic latitude. As can be inferred by Figure 2-right, the shape of the field lines is such to deflect more easily particles arriving at the geomagnetic equator allowing particles coming from the direction of the poles⁴

Inside the magnetosphere it is also possible to observe the phenomenon of particle trapping: there are two radiation belts, where protons and electrons are permanently trapped. The position of the two belts (the external is composed of electrons and the internal of protons) depends from energy of the particles. Human expeditions in space come across the part of the proton belt present in the South Atlantic region. Telecommunication satellites, usually placed in geostationary orbit at $6 R_{earth}$ are more sensible to electron belt (and transient effects related to solar events). For a discussion of the composition and structure of the radiation belts see ⁴.

⁴For this reason low energy $\approx 4 - 10 GeV$ cosmic rays balloon experiments need to fly at high latitudes (such as Canada). Satellites devoted to the study of solar and galactic cosmic ray component are usually launched in an high inclination (inclination close to 90° orbit).

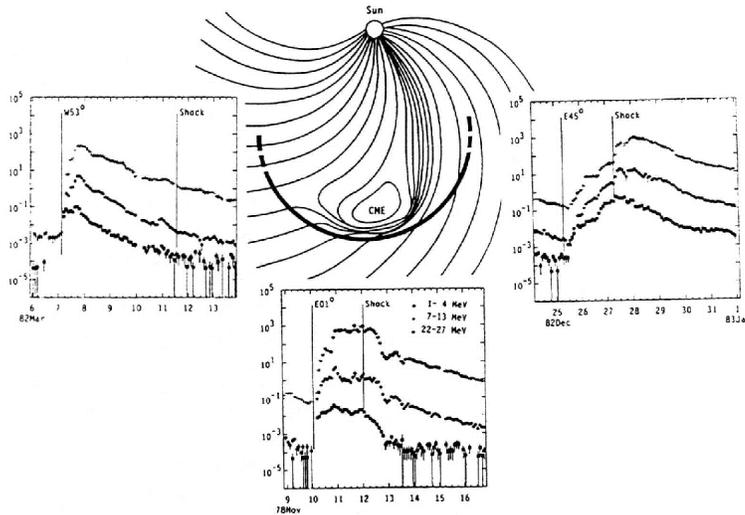


Figure 3: *Observation of a SEP from different heliospheric longitudes: the intensity vs time profile is sharper where there is direct magnetic connection to the emission region.*

1.3 Solar Energetic Particle Events

Solar Energetic Particle (SEP) events are the result of solar activity. They consist in a rapid particle flux increase (up to 4 orders of magnitude) emitted from active regions on the sun surface. It is currently believed that magnetic line recombination due to thermal motion of magnetic field frozen with the solar plasma can produce energy necessary for particle acceleration (typical released energies are $E = 10^{22} - 10^{25} J$ over a region of length $= 3 \times 10^7 m$ and height $= 2 \times 10^7 m$). This energies are released in a few seconds during solar flares, which consist an intense brightening (in visible and X-ray) of the solar surface, usually close to sunspots. It was therefore natural to believe that all particle acceleration took place at flare site. However recent studies are now moving toward a (blurred) division in two categories ³⁰):

- *Flare accelerated events*, where particle are accelerated at the flare site; they are usually small, impulsive, of low intensity and present an high ${}^3He/{}^4He$ ratio and heavy nuclei enhancement.

- *Coronal Mass Ejections*, where particle acceleration occurs at the shock front: they are of gradual nature and more intense, with little or no isotopic or nuclear increase over solar abundances.

This division is however somewhat artificial; there is agreement over the fact that there is no "typical" case and each event has its own peculiarities. This is also due to the fact that in addition to the different production and acceleration processes, there are also propagation effects in the interplanetary medium. In case of the smaller events, the kinetic energy of accelerated particles is small compared the solar wind kinetic and magnetic energy: particles therefore gyrate along the archimedean spiral of the interplanetary magnetic field as they leave the sun. These events can be observed only if the probe (or Earth itself) is located along the magnetic field line corresponding to the emission site. In case of CMEs, as shown in Figure 3, the kinetic energy of the emitted particles is intense enough to distort the magnetic field lines as the shock travels in the interplanetary space. CMEs can thus be observed over a much wider longitude range: as can be seen from Figure 3, the intensity vs time profile of the same event is very different according to the observation point: it is sharper in case of direct magnetic connection to the emission region and more gradual toward the boundaries.

1.4 Galactic Cosmic Rays

Galactic cosmic rays (GCR) are first accelerated in supernova explosions; in their propagation across the galaxy they undergo diffusion and shock acceleration processes. Despite the different production and interaction mechanisms they are subject to, the spectrum of cosmic rays above 1 *GeV* can be simply described by a power law of index -2.7 (see Figure 4-right). Cosmic rays extend to the highest energy so far observed: particles of $10^{20}eV$ have been detected⁷⁾. The dominant contribution to dose absorbed in space comes from GCR between $\approx 100MeV$ and $\approx 50GeV$: at lower energies (up to 1-2 *GeV*) GCRs are subject to solar modulation effects: their flux increases during periods of minimum solar activity and decreases during periods of maximum solar activity. Figure 5 shows the effect of solar modulation over the two last solar cycles; note how the shape of the two peaks (corresponding to solar minima) are different, a sign of the difference between the two solar polarities. From Figure 5 is also possible to see the presence of interplanetary gradients due to

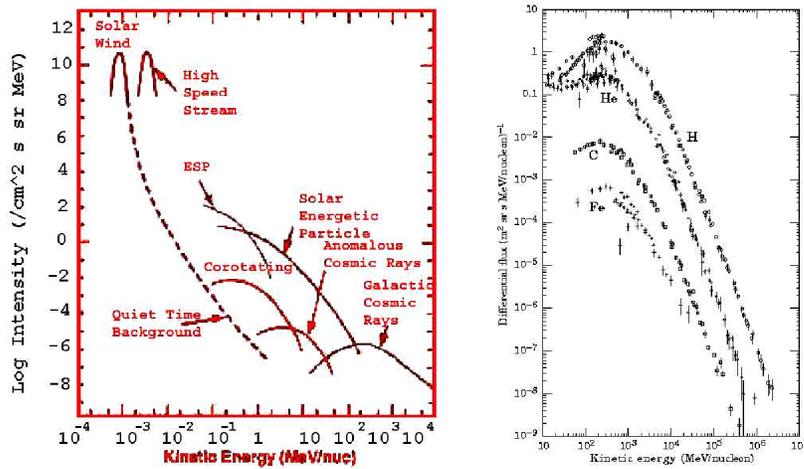


Figure 4: *Left: Solar cosmic ray differential energy spectrum for protons. The highest particle flux comes from the keV solar wind, with the more energetic ($\approx 10-100\text{MeV}$) events due to solar energetic particle events. Galactic cosmic ray differential energy spectrum for different nuclei (with a different scale on Y axis). Note the higher abundance of protons over heavier nuclei, the peak at 1 GeV below which solar modulation effects take place and the power law spectrum which extends to high energies ²⁶).*

solar wind: the comparison between IMP-7 and IMP-8 (located at 1 AU) with the Pioneer 10 and Voyager 2 probes shows that particle flux increases with distance from the Sun.

2 Radiation units of measure

Due to historical and biomedical reasons there are several different units to measure radiation: currently there is a gradual shift toward the use of SI units. Table 1 shows some of the different units used: the purpose is to correlate values related to activity (Becquerel) to the absorbed dose by living tissues (Gray) and to the damage radiation causes to tissues (Sievert). w_r represents the *radiation weighting factor*, which takes into account the fact that different particles may cause varying damage to tissues. Table 2 shows weighting factors for different particles: It is possible to see that, assigning a relative value to 1 to X, γ and electrons, alpha particles and nuclei have a w_r twenty times

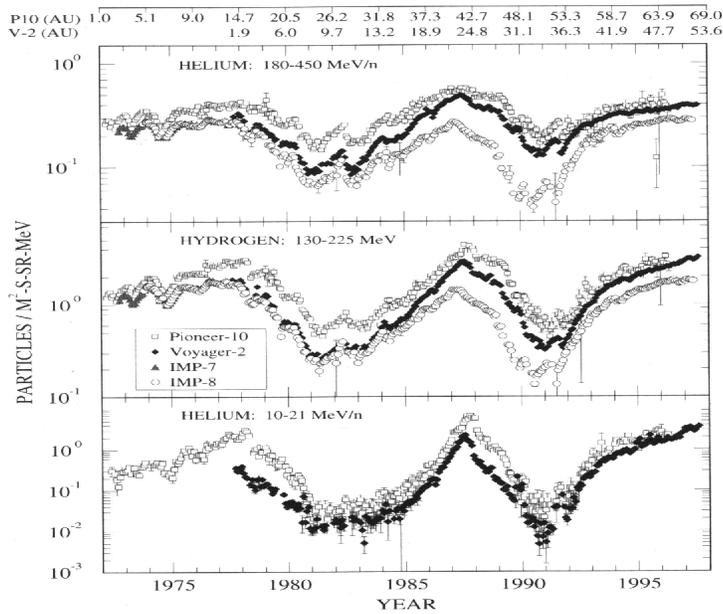


Figure 5: Solar modulation of galactic cosmic rays ²⁹⁾

higher due to the higher damage caused to tissues. In other terms, since the equivalent dose is in relation to the long-term risk of illness due to radiation, absorbing an equal dose of an alpha or an electron sources results in a risk of contracting cancer 20 times higher in the former case than in the latter.

The original relation between absorbed and equivalent dose was quantified with the introduction of the quality factor Q . In ICRP (International Committee of Radiation Protection) recommendations of 1990, "... the detail and precision inherent in using a formal Q - E relationship to modify absorbed dose to reflect the higher probability of detriment resulting from exposure to radiation components with high LET is not justified because of the uncertainties in the radiobiological information. In place of Q or more precisely \bar{Q} , the Commission now selects radiation weighting factors w_r , based on a review of the biological information, a variety of exposure circumstances and inspection of the results of traditional calculations of the ambient dose equivalent"

The use of w_r is particularly appropriate in presence of a single source (or a finite number of sources) typical of most work sites (laboratory sources, nuclear power plants, hospital, etc.). As we discussed, in space incoming radiation spans several orders of magnitude of energy and different particles, being composed of a non negligible part of heavy ions. Since in most cases it is not possible to identify the different contributions, a definition of Q depending

Table 1: Radiation units of measurement

Becquerel	unit of activity	1 Bq = 1 disint./s
Curie	<i>radioactivity of 1g of radium</i>	1 Ci = 3.7 10 ¹⁰ Bq
Gray	unit of absorbed dose Absorbed energy per unit mass	1 Gy = 1 Joule/kg
Rad	unit of absorbed dose	1 rad = 100 erg/g = 0.01 Gy
Unit of exposition	X or γ rad. in a point integ. over time. Ionizing charge per unit air volume	1 Coul/kg d'aria
Roentgen	Unit of exposition (Oldest unit of measurement)	1 R=2.1 10 ⁹ charge/cm ³ = 2.58 10 ⁻⁴ C/kg (STP)
Sievert	unit of equivalent dose	1 Sv=abs. dose(Gy)* w_r
Rem	Roentgen equivalent for man	1 rem = 0.01 Sv

Table 2: Relative Biological Effectiveness of radiation.

Particle type	w_r
γ, e^-, μ	1
Neutrons: E<10keV	5
10keV <E < 100 keV	10
100 keV <E < 2 MeV	20
2 MeV < E < 20 Mev	10
E>20 Mev	5
Protons E> 2MeV	5
α , fission fragment	20
heavy nuclei	20

from LET is used. Defining the Linear Energy Transfer (LET) as the energy deposited (in keV/ μm) in a given material by incoming particles. If $L(E)$ is the LET distribution function as a function of the energy deposited per unit distance E , the absorbed dose D is given by:

$$D = \int L(E)EdE \quad (4)$$

and the equivalent dose H is:

$$H = \int L(E)Q(E)EdE \quad (5)$$

Table 3: LET dependent quality factor.

Q	LET ($keV/\mu m$)
1	$E < 10keV/\mu m$
$0.32 \times L - 2.2$	$10keV/\mu m < L < 100keV/\mu m$
$300/\sqrt{L}$	$E > 100keV/\mu m$

where $Q(E)$ represents the energy dependent quality factor, defined in Table3 to take into account that damage to cells is constant up to a certain value, then grows linearly and finally decreases as the inverse of the square of E.

The lethal dose (defined as the dose resulting in 50% mortality in 30 days in absence of treatment) is equal to 2.5-3.0 Gy measured on the body longitudinal center line.

2.1 Radiation on Earth

Table 4: Typical doses on Earth.

Cosmic rays	0.26 mSv/yr 30nSv/hr
Pacemaker	1mSv/yr
Baggage check	0.02 μ Sv
Camping light	0.03 μ Sv
Concrete building	0.07 μ Sv
Computer	1 μ Sv/yr
50 km from carbon power plant	0.3 mSv/yr
50 km from nuclear power plant	0.09 mSv/yr

Typical background annual doses on Earth are 2.4 mSv (they vary between 0.4 mSv and 4 mSv according to different areas), of which 0.37 mSv are due to cosmic rays (at sea level). Other typical values from various sources are shown in Table 4. Radiation effects can be of deterministic and stochastic nature: the former (illness or death) appear in case of high doses, and the latter characterize low doses. In this case effects are evident as an increased death risk considering a statistically significant population of similarly exposed persons. However it is very difficult for the very low doses typically absorbed to identify the role and importance of different causes such as pollution or smoke. CERN recommended limits to exposure of radiation is 15 mSv/year; in the USA is 50 mSv/year.

2.2 Radiation at high altitudes

Comparing radiation limits with Table 4 is clear that safety thresholds and typical background contributions are well below any dangerous levels: the contribution due to cosmic rays is negligible thanks to the geomagnetic and atmospheric shielding. At higher altitudes the atmospheric shielding is greatly reduced, in addition interaction of high energy cosmic rays in the atmosphere produces electromagnetic and hadronic showers which increase the particle flux. In Figure 6 is shown the intensity of charged particles as a function of altitude. The highest intensity is found at altitudes of 10 km (the altitude of intercontinental flights) where showers reach a peak before being reabsorbed. The contribution of cosmic rays to the dose absorbed is such to make them being considered a working health hazard by airline crew and frequent flyers. A number of measurements has been carried out in order to estimate the nature, energy composition and - most importantly - the dose of cosmic rays during airplane flights. Some of the doses measured during these flights are shown in Table 5 and show how the occupational dose is higher than average world population but well below any safety threshold. However this value may increase in extremely particular conditions such as flight during an intense Solar Particle Event, especially at high geomagnetic latitudes.

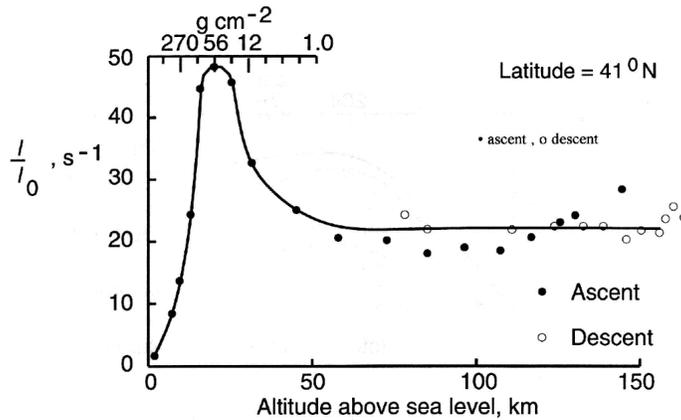


Figure 6: Particle intensity as a function of altitude ⁹⁾

2.3 Radiation in Space

In space, the absence of atmosphere becomes even more significant: astronauts and equipment may be subject - according to the orbit - to intense fluxes of radiation. Almost all manned flights took and take place in Low Earth Orbit

Table 5: Absorbed doses due to cosmic rays during airplane flights ³²⁾.

Route	Duration (min)	Dose (μSv)
Dublin-Paris	95	4.5
Dublin-Rome	180	10
Stockholm-Tokyo	605	51
Amsterdam-Vancouver	645	70
Occupational	1 year	1000-6000

(LEO), at altitudes varying between 200 and 500 km, with the only notable exception of the Apollo program, which not only took man on the Moon but also outside Earth's geomagnetic shielding. Most LEO missions (such as Mir, the International Space Station and Shuttle) have a low inclination (51.6°) to avoid high latitude areas, where the lower geomagnetic cutoff results in an higher cosmic ray flux. Cosmic rays and radiation measurements inside spacecraft have been the subject of intense investigation throughout the course of space exploration. These studies are particularly difficult since they need to take into account the orbit dependence of cosmic ray flux and its propagation inside the varying absorber thicknesses of the spacecraft. Investigations therefore include dosimetric measurement (often with different absorber thickness or inside human phantoms to simulate propagation inside the body ²⁷⁾) which produce information on the total absorbed dose. A number of more sophisticated detectors are also used to monitor in real time long and short term variations in space ^{18, 19)}. Figure 7 shows the particle flux measured on board Shuttle (at an altitude of 300-400 km) with the detector RRMD-III, as a function of the point of the orbit. It is possible to see how the most significant contribution to dose in space comes from trapped protons present in the South Atlantic Anomaly. To estimate the total dose it is necessary to integrate the differential LET (Figure 8) curves according to eq. 4 and 5; the resulting value amounts to 0.516 mGy/day (with an effective quality factor $Q=1.81$ this amounts to 0.935 mSv/day). Of this the absorbed dose rates are $0.120 \mu Gy/min$ for the galactic component and $4.80 \mu Gy/min$ for the trapped proton component. Doses in orbit are therefore much higher than on ground: assuming these values equal for Mir or the ISS, this amounts to 188 mGy/year. This value can be reached for long permanence flights such as those on Mir and can be even higher due to solar events contribution.

In addition to orbit dependent dose differences it is possible to study the differences of nuclear composition in different points of the orbit ⁸⁾ to show latitude dependent effects. Mir orbit was divided according to the McIlwain

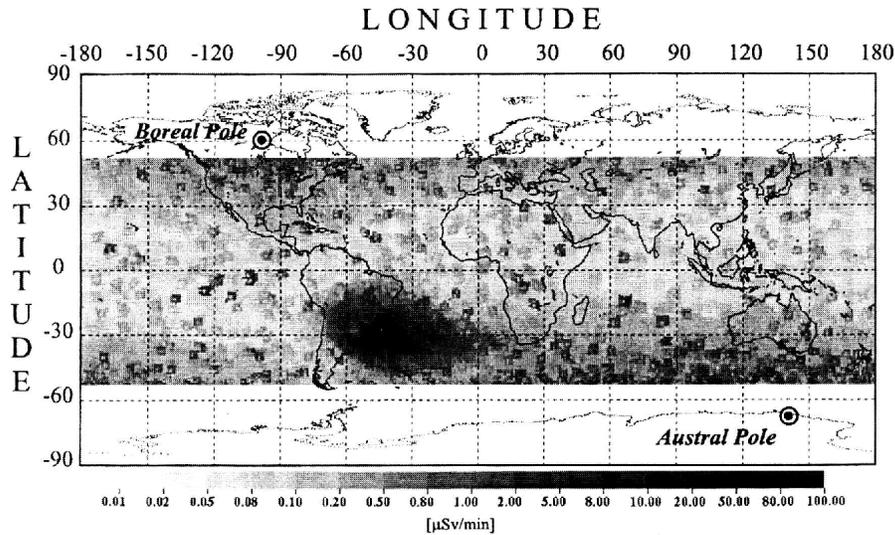


Figure 7: Dose rate on board Shuttle 19)

parameter⁵ L and the geomagnetic field B in three regions: Galactic Cosmic Ray region (GCR, $L > 2$), South Atlantic Anomaly (SAA, $L < 2$ and geomagnetic field $B < 0.25\text{G}$), and the remaining region ($L < 2$, $B \geq 0.25\text{G}$). This reference system is particularly useful since charged particles spiral along the magnetic field and bounce between the mirror points at values of constant L . In a given point of the orbit, the geomagnetic cutoff C determines the minimum energy for primary cosmic rays to reach Mir and be detected. Note that this value is valid for particles orthogonal to the local field line and outside Mir. In addition, particle energy inside the station can be modified by the interposed material of the station and the presence of nuclear interactions, so it should be used only as a reference. Particles with energy above is 90 MeV/n (of which at least 20 MeV/n are lost in the Al of the station) are being considered. Naturally the 3mm Al thickness assumed only represents a lower value, since different amounts of the station and the equipment contained can be interposed between the detector and the local field line along which the particles come.

⁵The McIlwain parameter L represents - at a first approximation - the value (expressed in Earth radii) at which the magnetic field line passing through the point considered intersects the geomagnetic equator. For LEO such as Mir values close to $L = 1$ are in proximity to the equator and increase at higher latitudes. For a detailed definition see 16).

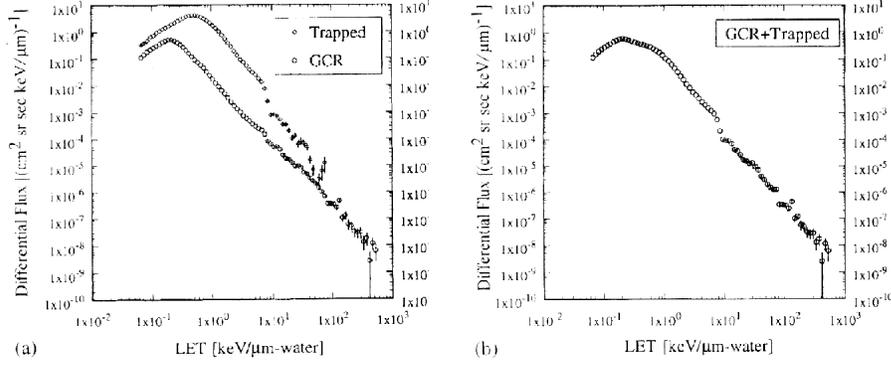


Figure 8: *LET distribution measured on board shuttle ¹⁹*. Note how the trapped component (higher curve of left panel) is dominant for low LETs

Table 6: Relative abundances normalized to carbon in the three regions for particle with $E > E_{con}$ (see text).

Z	$L > 2$ ($C > 0.6 GV$)	$L \leq 2$ ($C > 3.9 GV$)	SAA ($C > 3.9 GV$)	600-1000 MeV/n ²⁶)
5 (B)	0.63 ± 0.09	0.53 ± 0.35	0.55 ± 0.09	0.307 ± 0.005
6 (C)	1 ± 0.1	1 ± 0.06	1 ± 0.12	1 ± 0.02
7 (N)	0.41 ± 0.06	0.34 ± 0.08	0.22 ± 0.04	0.274 ± 0.007
8 (O)	0.65 ± 0.07	0.66 ± 0.08	0.77 ± 0.17	0.93 ± 0.02
10 (Ne)	0.33 ± 0.06	0.13 ± 0.02	0.12 ± 0.02	0.149 ± 0.004
12 (Mg)	0.07 ± 0.02	0.13 ± 0.02	0.12 ± 0.02	0.187 ± 0.005
14 (Si)	0.05 ± 0.02	0.1 ± 0.02	0.1 ± 0.02	$.13158 \pm 3 \cdot 10^{-5}$

At $L = 2$, $C = 3.9 GV$ while at high latitude ($L=4.4$) $C = 0.8 GV$. These two values represent the minimum cutoff for a given region; they correspond to a minimum kinetic energy (for particle with mass/charge ratio of 2) of $\approx 150 MeV/n$ ($C = 0.8 GV$) and $\approx 1600 MeV/n$ ($C = 3.9 GV$). At this energies particles lose only a small fraction of their kinetic energy in crossing the hull of the station: again, using 3 mm of Al as a reference, we find that, for instance, a 150 (1600) MeV proton loses 3.4 (1.3) MeV to enter the station. In case of other nuclei, the values are similar: if we consider a carbon nucleus, we have 4.8 (1.2) MeV lost for 150 (1600) MeV/n. The particle distributions of the three regions are shown in Figure 9. The continuous line shows the galactic component which has an higher flux due to the lower geomagnetic cutoff. This

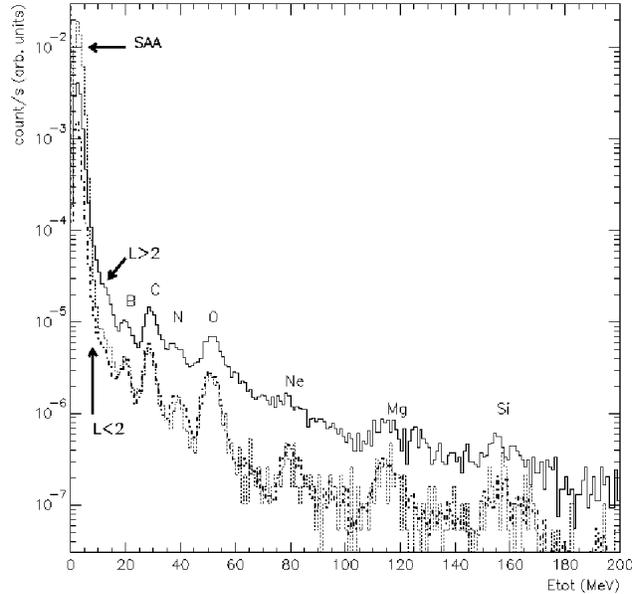


Figure 9: High energy nuclear abundances inside Mir: Continuous line: galactic ($L > 2$) component; Dotted line: SAA component ($L < 2$, $B < 0.25G$); Dashed line: remaining region ($L < 2$, $B \geq 0.25G$). The SAA region has a higher proton flux due to trapped particles but $Z > 5$ particles are equally abundant to the $L < 2$ region due to the equivalent cutoff. In case of the galactic component, the lower geomagnetic cutoff results in a higher integral particle flux so that $Z > 5$ nuclei are more abundant ⁸⁾.

allows particles with lower energy to reach Mir resulting in a higher particle count. The wider energy range implies a larger energy release range, resulting in the peaks to be less sharply defined. In this range, proton and helium flux is lower than that measured in the SAA (dotted line) where the trapped component is dominant if compared to galactic and $L < 2$ abundances. Indeed the $L < 2$ curve (dashed) has a lower $Z \leq 2$ flux if compared to SAA but an equal $Z \geq 5$ flux, since in both cases the component selected at this energy is the same. From these distributions it is possible to reconstruct relative abundances for the different nuclear species (shown in Table 6). Table 6 also shows the relative cosmic ray abundances at 1AU ²⁶⁾ measured in the energy range of $\approx 1GeV$. It is possible to see how, especially for the $L < 2$ regions, notwithstanding the bulk of the Mir, the data are in general in agreement with the following differences:

- An overabundance of B in respect to C. It is roughly twice the 1AU value in all three regions. This could be accounted as secondary production

due to hadronic interactions.

- A higher amount of N in the $L > 2$ region compared to the other regions and 1AU data. This could be due to an larger production of secondary N at lower energies.
- A lower amount of Oxygen nuclei in $L > 2$ and $L < 2$ regions (SAA value is in agreement within errors with 1 AU data). Also in this case the effect can be explained with an higher hadronic interaction cross section for O in respect to C: Oxygen could be considered as composed of four alpha particles (He nuclei) and Carbon of three. Thus, if we assume the ratio of the cross sections to be equal of the ratio of the nucleons of Carbon and Oxygen ($12/16 = 0.75$) and we multiply by the original flux of 0.93 we obtain an abundance of 0.7.
- An higher amount of Ne and lower amount of Mg and SI in the $L > 2$ region.

In all cases it is clear that a crucial role is played by hadronic interactions in the matter, but an accurate estimate of the processes involved is complicated by the estimation of the energy-dependent cross sections and the amount of material interposed between the detector and the exterior of the station.

2.4 Other effects of radiation: the Light Flash Phenomenon

The study of nuclear abundances inside manned spacecraft is of particular relevance for its relationship with other radiation related effects, which may be symptoms of different phenomena due to exposure of cosmic rays. The most notable is the so-called "Light Flash effect": Light Flashes (LF) are visual phenomena originated by the interaction of cosmic radiation with the human eye. This phenomenon was predicted by Tobias in 1952¹²⁾ and first observed on board Apollo 11. Subsequently, a number of investigations in space (on board Apollo, Apollo-Soyuz Test Project - ASTP, Skylab, Mir^{13, 14)}) were performed, showing different LF shapes and topologies. Evidence tends to ascertain the effect to the interaction of cosmic rays with the eye, with heavy ($Z \geq 6$) nuclei having an higher probability of causing this effect. A number of ground experiments was also performed using accelerator beams and resulting in the observation of this effect with different particles. Several hypothesis were put forward to explain LFs: they included the generation of Cherenkov light in the vitreous, knock-on electrons or direct ionization in the retina. Although a ground experiment which scanned different regions of the eye with a nitrogen beam seemed to pinpoint the LF phenomenon to occur in the posterior globe of the eye, several of the aforementioned causes may be possible.

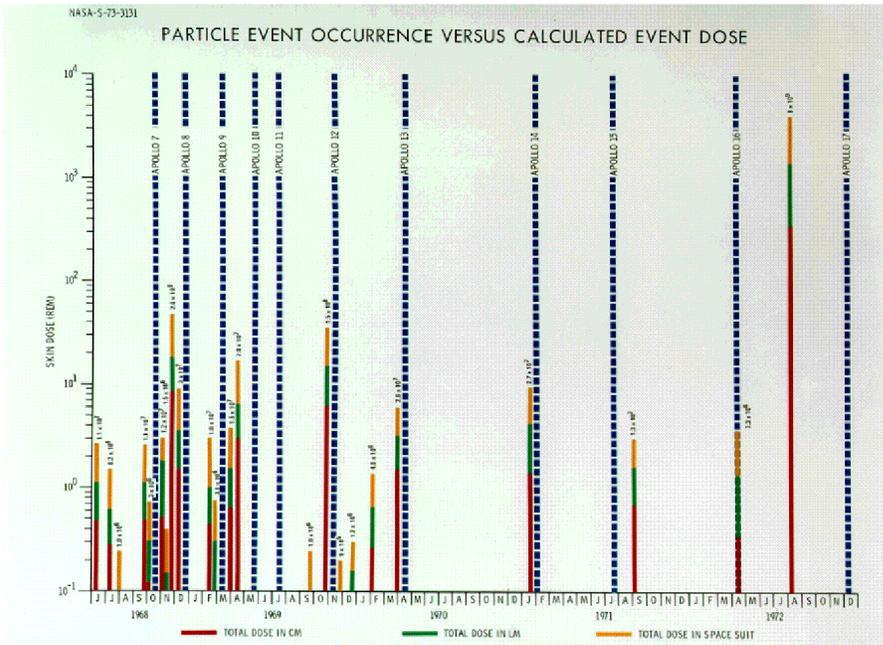


Figure 10: *SEP event occurrences and intensities compared with Apollo missions 31).*

2.5 Interplanetary missions

As we already mentioned, man ventured outside the geomagnetic shielding of the Earth only during the Apollo missions. In this case the radiation risk due to SEP events is greatly enhanced 28): Apollo astronauts were shielded only by the spacecraft; on the Moon, which has no magnetic field, the protection was even lower, especially when performing activities outside the LEM. Fortunately no such event occurred, event though (Figure 10) several very intense events (such as that between Apollo 16 and 17 missions) occurred between missions which could have jeopardized the mission and even killed the astronauts.

If these risks and some countermeasures to protect the astronauts could be taken in the short ($\simeq 2$ weeks) Apollo missions, a mission to Mars - which would last more than two years - would imply almost certain exposition of the crew to several SEP events, some of which of high intensity.

2.6 Passive and Active shielding

Crew protection of astronauts during missions can be achieved with passive or active ways. In the first case is necessary to build at least a part of the spacecraft with higher amount of shielding. The disadvantage of such a system

relies in the inevitable trade-off with total weight of the hull: increasing the thickness of the hull from 2 to 9 mm for a typical module would increase the cutoff energy of solar particles from 20 MeV to 60 MeV but would imply a weight increase of 1.5 tons.

An alternative approach is to use an *active* system^{24, 25}, composed by a magnetic lens which would act in a way analogous to Earth's magnetic field, deflecting particles outside the habitable area. This lens would be composed by a superconducting solenoid which could be switched on - given sufficient warning by a satellite monitoring network. Such a network would require use of systems observing the sun surface, corona and analysis of each event to determine - amongst other things - the nature, intensity and magnetic connection between the SEP and the spacecraft, in order to turn the solenoid in the direction of the incoming particles.

2.7 Other effects

In this work we have covered briefly the *direct* effects of solar and cosmic rays, related to the ionizing nature of radiation. In reality, in the solar terrestrial environment there are many *indirect* effects related to the interaction between solar particles and our magnetic field.

There are many adverse effects: especially at high geomagnetic latitudes, currents induced in electric power lines by the magnetic field generated by high altitude currents (*electrojet currents* - in turn induced by Solar particle events) have caused in many occasions power failures. In electric lines voltage differences of up to 6 kV/km have been observed: these highly fluctuating power spikes often short out power generators (one of the most intense events caused a blackout in Quebec in 1989).

Other similar effects - also on transatlantic telephone cables have been observed during maximum solar activity (such as in 1972 and 1958); back in 1840 and 1880 the telegraph system in USA could be operated on several occasions without batteries. Another long term problem is the oil duct corrosion (again due to currents induced in pipes by electrojet currents) which - in addition to reducing the life span of these systems - was also the cause in 1989 of a train accident which - upon passage - ignited the gas which was leaking from the duct.

3 Conclusions

In this work we have tried to show the complex interrelations between the Sun, our galaxy and the Earth environment focusing mainly on the radiation effects. The study of the solar terrestrial environment as a whole is a rapidly growing

interdisciplinary field of interest not only for manned exploration of space but for an understanding of our planet and its relationship with the solar system. There is evidence^{33, 34)} that we are reaching the end of a Gleissberg (85-90 year) cycle of the Sun, which would imply that future Solar cycles could be characterized by more frequent SEP. What would be the impact of this higher solar particle rate on human activities both in space and on Earth is still to be understood.

References

1. M. Longair, 1992, *High Energy Astrophysics* Vol. 1 & 2, Cambridge University press.
2. M. G. Kivelson and C. T. Russel eds, 1996, *Introduction to Space Physics*, Cambridge University press.
3. J. K. Hargreaves, 1995, *The solar-terrestrial environment*, Cambridge University press.
4. W. Hess, 1965, *Introduction to Space Science*, Gordon and Breach Science Publishers.
5. W. Larson, and J. Wertz, 1992, *Space mission analysis and design*, Kluwer academic Publishers.
6. F. L. Deubner, D. O. Gough, (1991), *Ann. Rev. Astron. Astrophys*, **22**, 593.
7. P. K. F. Grieder, 2001, *Cosmic Rays at Earth*, Elsevier.
8. V. Bidoli, M. Casolino et al., 2001, *J. Phys. G*, **27** 2051
9. Schaefer, H. J., 1979, *Radiation and man in space*, Adv. Space Science 1, 267.
10. Selesnick R S, Cummings A C, Cummings J R, Mewaldt R A, Stone E C and von Rosenvinge T T 1995 *J. Geophys Res.* **100** (A6) 9503
11. A. J. Lazarus, J. W. Belcher, K. I. Paularena, J. D. Richardson, 1998, *Space Scie. Rev.*, **83**, 87.
12. C. A. Tobias, 1952, *J. Aviat. Med.* **23** 345
13. F. J. D'arcy, N. A. Parker 1962 *Nature* **196** 1013

14. M. J. Malachowski, (1978) LBL Report LBL-5683, National Technical Information Service, Springfield, Virginia
15. G. Horneck 1992 *Nucl. Tracks Radiat. Meas.* **20** 1 185
16. C. E. McIlwain, 1961, *JGR* **66**, 3681
17. P. J. McNulty 1996 *IEEE Trans. on Nucl. Sci.* **43** 2 475
18. G. Reitz *et al.* 1996 *Radiat. Meas.* **26** 6 679
19. T. Sakaguchi, T. Doke, *et al.*, 1999 *NIM A* **437** 75
20. V. Bidoli, M. Casolino *et al.*, (2000) *Adv. Space Res.* **10** 2075.
21. S. Avdeev, *et al.*, *submitted to Acta Astronautica.*
22. V. Bidoli, M. Casolino, *et al.*, 2001 *ApJS* **132** 2 365
23. R. Sparvoli, V. Bidoli, *et al.*, 2000 *Nuclear Physics B (Proc. Suppl.)* **85** 28
24. P. Spillantini 2000 *Nuc. Pys. B (proc. suppl.)* **85** 3
25. P. Spillantini, F. Taccetti, P. Papini, L. Rossi, M. Casolino, 2001, *Physica Medica*, 17 supp. 1.
26. J. A. Simpson, 1983 *Ann. Rev. Nucl. Part. Sci.*, **33**, 323.
27. D. Badhwar, F. A. Cucinotta, 1998 *Rad. Res.*, **149**, 209
28. Letaw, J. R., Silberberg, R., and Tsao, C. H., 1987, *Nature* 330, 709-170.
29. McDonald, F., 1998, *Sp. Sci. Rev.*, 83, 33.
30. Reames, D., 1999, *Sp. Sci. Rev.* 90, 413.
31. NASA S-73-3131
32. European Commission, 1997, *Recommendations for the implementation of Title VII of the European Basic Safety Standard Directive concerning significant increase in exposure due to natural radiation sources*, Luxembourg.
33. K. G. McCracken, G. A. M. Dreschhoff, D. F. Smart, and M. A. Shea, 2001, *Proc. of ICRC*, SH1.4 35, 3205.
34. K. G. McCracken, D. F. Smart, M. A. Shea, and G. A. M. Dreschhoff, 2001, *Proc. of ICRC*, SH1.4 36, 3209.

Protons and leptons trapped in the Earth's radiation belts

G. Esposito, E. Fiandrini
University and INFN of Perugia, Italy

ABSTRACT

Accurate measurements of lepton fluxes in the energy range 0.2–10 GeV have been performed with the Alpha Magnetic Spectrometer (AMS). We also analyzed proton fluxes at rigidities above 400 MV. The flux distributions as a function of the canonical adiabatic variables L , α_o are presented for the interval $0.95 < L < 3$, $0^\circ < \alpha_o < 90^\circ$ for electrons ($E < 10$ GeV), positrons ($E < 3$ GeV) and protons ($R < 10$ GV). The results are compared with existing data at lower rigidities and in similar (L, α_o) range.

1 Introduction

For the low altitude (250-1000 Km) region near the Earth, both for commercial and scientific purpose, it is particularly important to accurately model the radiation environment. ¹⁾ proved the existence of ~ 100 MeV trapped leptons both in the Inner Van Allen Belts (*stably-trapped*) and in the region below (*quasi-trapped*), and determined their charge composition. No accurate measurements were available for the composition and flux intensity in the near Earth region

in the AMS rigidity range until the flight in 1998. In the following, we use the high statistics collected by the AMS experiment in 1998 to present a detailed study of under-cutoff fluxes in the O(10 GV) rigidity region. They are analyzed in terms of the canonical invariant coordinates of the particles motion, the L parameter, the equatorial pitch angle with \vec{B} field, α_0 , and the mirror field B_m .

2 AMS and the STS-91 flight

AMS is large acceptance ($\sim 0.16 \text{ m}^2 \text{ sr}$) detector designed to operate on the International Space Station for three years. A prototype version of the detector, equipped with a high resolution silicon tracker embedded in a permanent magnet, a plastic scintillator Time-Of-flight system, and a threshold Aerogel Cerenkov counter for ep discrimination, operated onboard the shuttle Discovery during a 10-day in June 1998. It orbited at altitudes between 370-390 km with an inclination of 51.7° in the GTOD reference frame ²⁾. The SAA region was excluded in this analysis.

3 Under-cutoff lepton and proton classification

To reject the cosmic component of the measured lepton fluxes, the particle trajectories in the Earth's magnetic field were traced. A particle was classified as *quasi trapped* if its trajectory reached an altitude of 40 km, taken as the dense atmosphere limit, before or after its detection in AMS. To avoid Earth penumbra an effective cutoff R_{eff} was defined as the maximum rigidity value at a given magnetic latitude θ_m for which no traced lepton was found to be of cosmic origin. The total time spent by each particle above the atmosphere, before and after detection, as function of energy is shown in Fig. 1 for positrons; an identical behaviour is observed for electrons. The geographical location where the trajectories intercept the atmosphere determine the leptons *production* and *impact* points is shown in Fig. 2 The lepton residence times do not exceed ~ 30 s, with 52% of e^- and 38% of e^+ having a $T_f < 0.3$ s independent of their energy, referred to as the *short-lived* populations in ²⁾. The corresponding impact/production points for the latter are spread over two bands on either side the magnetic equator shown as yellows in Fig. 2. A scaling law, $T_f \approx E^{-1}$, is observed for the remaining leptons (the *long-lived* population), localized in well defined spots with sinks for positives being the source for negatives. A sample of protons detected in equatorial regions ($\theta_m \leq 17^\circ$) were traced. The same features as for positrons were observed both for residence times and impact/production points. It is not practical to trace each proton due to the large statistical sample O(10^6). We defined an effective cutoff R_{eff} using a modified Stormer cutoff and taking into account the rigidity resolution of the detector. The behaviour observed in the under-cutoff lepton and proton data

can be explained in terms of the geometry of the *drift shells* crossed by the AMS during the shuttle flight, and in particular by the fact that all the shells evolve partially in the atmosphere. Therefore no permanent trapping can occur. The residence times are determined by the periodicity of the drifting (T_d) or bouncing ($T_b \ll T_d$) motion, and depending on the fraction of the mirror points below atmosphere. The impact/production points correspond to the intersection of the shell surfaces with the atmosphere, where particles generated in interactions are injected into the shells ⁴⁾.

4 Under cutoff flux

For the description of under cutoff fluxes, the energy E (the rigidity R for protons), the L-shell and the equatorial pitch angle α_0 were used. Two different maps for two different energy bins for e^+ and e^- are shown in Fig. 3. The flux is limited by the cutoff rigidity R_c : on a given shell only particles with $R \leq R_c$ are allowed to populate the shell. The e^+ , e^- flux maps ratio in the energy interval 0.2-2.7 GeV is shown in Figs. 4 where the solid line identifies the boundary in (L, α_0) for which no leptons can be found with $T_f > 0.3s$. Above the curve, at higher values of α_0 , the long-lived component of fluxes begins to dominate. Above $\alpha_0 > 60^\circ$ the flux is substantially due to the long-lived component; here the e^+ intensity represents $\approx 80\%$ of the total leptonic flux. In L, the long-lived component dominates only at very low values where the positron excess is more pronounced. The energy spectra for particles with $\alpha_0 \geq 70^\circ$, shown in Fig. 5, with superimposed measurements from MARIYA at lower energy ¹⁾; the e^+ is higher than the e^- flux by a factor ~ 4.5 , in contrast with MARIYA data which indicate the same level of flux for both e^+ and e^- . The differential flux maps for protons in four rigidity bins are shown in Fig. 7 with an observed intensity smaller than $1 [m^2 s sr MV]^{-1}$, well below the expected intensity for a trapped component. A rapid decay of the intensity is observed as the rigidity increases. The intrinsic azimuthal β angle allows a separation between local magnetic East and West (particles coming from West have always $\beta < 0$, while those coming from East $\beta > 0$). Integral fluxes for $\beta < 0$ and $\beta > 0$ in two rigidity intervals ($.37 \leq R \leq 3.7$ GV, $3.7 \leq R \leq 10$ GV), integrated over α_0 and L are shown in Fig. 8 A) and B), respectively. The East-West flux asymmetry A, defined as $(J_{\beta < 0} - J_{\beta > 0}) / (J_{\beta < 0} + J_{\beta > 0})$, as function of L and α_0 are also shown in Fig. 8 C) and D), respectively. At low altitudes the proton flux appear to be strongly anisotropic due to the gyroradius and geomagnetic cutoff effect ⁵⁾. In Fig. 6 a comparison between AMS proton fluxes and trapped AP8 for locally mirroring particle ($\alpha_0 \geq 70$, $L = 1.2$) is shown. In ⁶⁾ was used a complete MC to reproduce the AMS under cutoff spectra where the albedo component is completely explained by secondary production from interactions at the intersections of drift shells with atmosphere. The spectra obtained from

secondary production ⁶⁾, shown in Fig. 6 superimposed to data, well agree with the observed flux.

5 Conclusion

The AMS data clearly indicate the existence of protons and leptons underneath the Inner Van Allen belts for particle energies of several GeV. The measured fluxes are not stably trapped since the corresponding drift shells are not closed over the SAA region. Charge composition shows a dominance of positively charged leptons in a definite region of the (L, α_o) . Strong East-West anisotropies are observed for protons at rigidities above 3.7 GV. The interaction of primary cosmic rays and inner radiation belt protons with atmospheric nuclei naturally leads to a e^+ excess over e^- and seems suitable to explain the observed charge ratio for *quasi-trapped* flux.

References

1. A.M. Galper et al., Geophys. Monogr., 1997.
2. J. Alcaraz et al., Phys. Lett. B, 484, p.10, 2000, 472, p.215, 2000
3. L. Derome et al., Phys. Lett. B, 489, p.1-8, 2000.
4. Fiandrini et al., Proc. of 27th ICRC, 2001
5. Esposito et al., Proc. of 27th ICRC, 2001
6. Zuccon et al., Proc. of 27th ICRC, 2001

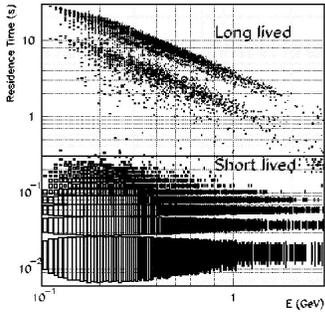


Figure 1: Residence time vs kinetic energy for e^+ .

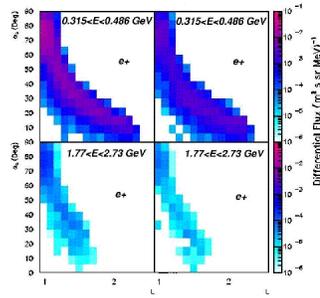
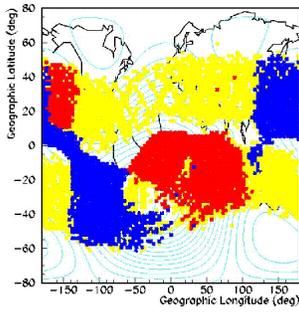


Figure 2: Geographical distribution of Figure 3: e^+/e^- differential flux maps prod/impact points. Yellow short lived, for 2 different energy bins red/blue bands long lived positrons.

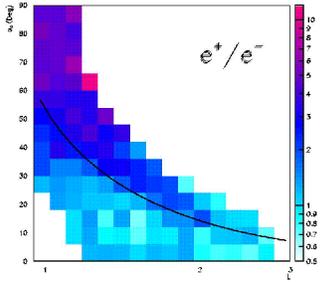


Figure 4: Integral e^+/e^- ratio between $0.205 \leq E \leq 2.73$ GeV

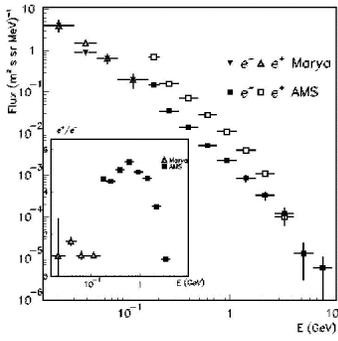


Figure 5: Energy spectrum comparison between AMS and Marya for e^+ and e^- $\alpha_0 \geq 70^\circ$ and $L=1.2$. In the small plot the e^+/e^- ratio.

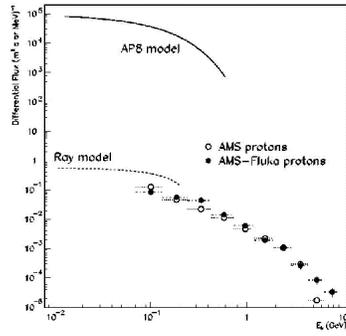


Figure 6: Energy spectrum comparison between AMS, AP8 and MC data for particles with $\alpha_0 \geq 70^\circ$ and $L=1.2$.

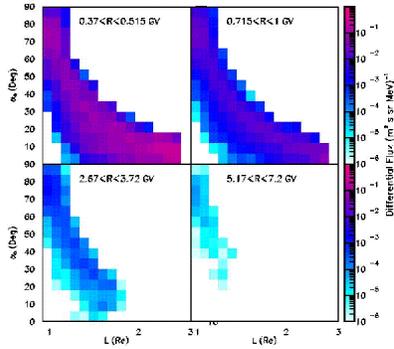


Figure 7: Differential flux maps for 4 different rigidity bins

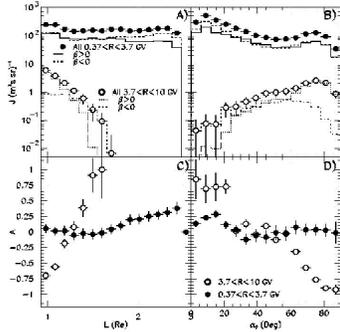


Figure 8: Integral flux and asymmetries as function of α_0 (B,D) and of L (A,C) for protons

COSMIC RAYS IN THE EARTH VICINITY

Roberta Sparvoli

University of Rome "Tor Vergata" and INFN section of Roma2

ABSTRACT

This article presents a description of the radiation environment around the Earth, with special emphasis on how the Earth's magnetic field and the Sun act on the flux of primary cosmic rays.

1 Introduction

Observing cosmic rays from Earth is like making astronomical observations from the bottom of an ocean. In such a situation the constant swirling of the ocean surface would distort all images. Moreover, the observer would have to correct for the diffraction of light as it passes from the air into the water, if he wants to reconstruct the true direction. Finally, an occasional oil spill might absorb part of the light, so that finally the resulting picture of the astronomical sky would certainly be distorted.

Cosmic rays arriving at Earth are survivor of primary cosmic rays or are a product of them. They have survived collisions with gas atoms in the interstellar and interplanetary space, and may have collided in atmosphere. Raw observations thus provide only the first stage in reconstructing the cosmic rays journey from the production site to Earth.

Moreover, just as important as collision, is the effect of the Earth magnetic field and that of the Sun, and this is specifically what will be illustrated in this article 1, 2).

2 The Earth's magnetic field

The Earth's magnetic field is generated by electric currents flowing deep inside the solid Earth 3). To a first approximation, at and close to the planet's surface it can be represented as a *dipole field*, whose poles are at geographic latitude and longitude 79°N, 70°W, and 79°S, 110°E, reversed with respect to geographic poles, about 11° inclined with the Earth axis and shifted by 320 km. The field changes slowly over the years, producing a secular drift of the magnetic poles.

The magnetic flux density is given by:

$$B(r, \lambda) = \frac{M}{r^3}(1 + 3\sin^2\lambda)^{\frac{1}{2}}, \quad (1)$$

where M is the dipole moment, r the geocentric radial distance and λ the magnetic latitude (see figure 1). The field components in the r and λ directions are:

$$B_r = -2M\sin\lambda/r^3, \quad (2)$$

$$B_\lambda = M\cos\lambda/r^3. \quad (3)$$

Thus

$$\frac{B_\lambda}{B_r} = \frac{rd\lambda}{dr} = -\frac{1}{2\tan\lambda}$$

and, by integration,

$$r = r_0\cos^2\lambda. \quad (4)$$

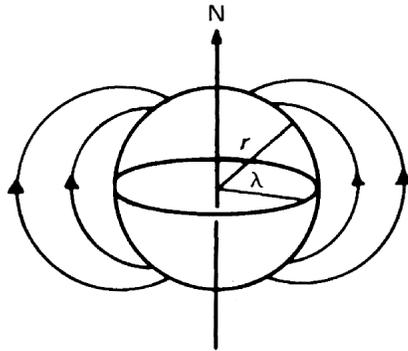


Figure 1: *The geomagnetic latitude λ relative to the magnetic north pole.*

For a selected value of r_0 this is the equation of a field-line, the locus of the force acting on a single north pole. If $\lambda = 0$, $r = r_0$; r_0 is thus the radial distance to the field-line over the equator, and its greatest distance from the Earth.

It is convenient to use the radius of the Earth, R_E , as the unit of distance. Putting $r/R_E = R$,

$$B(R, \lambda) = \frac{0.31}{R^3} (1 + 3\sin^2 \lambda)^{\frac{1}{2}} \text{ gauss}, \quad (5)$$

0.31 G ($=3.1 \times 10^{-5}$ Wb/m²) being the flux density at the Earth's surface at the magnetic equator. In these terms, the field-line equation becomes

$$R = R_0 \cos^2 \lambda, \quad (6)$$

R and R_0 being measured in Earth-radii. The latitude where the field-line intersects the Earth's surface is given by

$$\cos \lambda_E = R_0^{-\frac{1}{2}}. \quad (7)$$

The dipole representation is accurate to about 30% at distances $\leq 2-3 R_E$. A better empirical representation is based on a multipole expansion (International Geomagnetic Reference Field IGRF model), with slowly time-dependent coefficients.

To describe the field, also in non-dipole approximation, usually the McIlwain coordinates (B, L) are used. Taken a point P on the Earth's surface, we

consider the flux line through P and we measure its distance to the Earth's center on the equatorial plane. Such distance, in units of Earth-radii, is the L coordinate; with $L - shell$ we indicate the shell of points with the same value of L , while B is given by the value of the scalar magnetic field in the point. In this way, a couple of coordinates (B, L) is associated univocally to every point in the space.

2.1 Particle trajectories

The paths followed by cosmic rays as they approach the Earth are influenced by the Earth's magnetic field. Several factors control the path of a cosmic ray (CR): its initial direction, its mass speed and electric charge, and how the magnetic field varies with distance to the Earth. As a result, the direction of motion changes continuously as a CR approaches the Earth, resulting in a highly convoluted path.

The influence of the Earth's magnetic field on cosmic rays was first seen through what was called the "*latitude effect*". Around 1930 experiments conducted at ground and on the mountains showed that the flux of cosmic rays reaching detectors had its minimum at the magnetic equator while was increasing at northern latitudes. Further investigations confirmed this effect, which was explained by the fact that charged cosmic rays arrive randomly from all directions before they encounter the Earth's field, and begin to be deflected. Only the fastest, most energetic particles can penetrate to the region of the magnetic equator. Less energetic particles are detectable at mid-latitudes, and there are no restrictions on the energy of the detectable primary particles. The latitude effect showed clearly that cosmic rays are *charged*.

If we inject particles with the same value of pc/ze into a magnetic field \mathbf{B} at the same *pitch-angle* α (angle between the velocity vector and the magnetic field direction), they have exactly the same dynamic behaviour. The quantity pc/ze is called the **magnetic rigidity** of the particle. It has dimensions of volts; a useful unit for practical purposes is gigavolts (GV). Developing the calculations of charged particles in a magnetic field it can be seen that at each point on the Earth's surface, there is a threshold energy, called *geomagnetic cut-off*, below which the particles cannot reach the surface: the geomagnetic field acts as a *filter* for the particles. The geomagnetic access criterion is given by the well-known approximation first derived by Fermi

$$R = (A/Q)(E^2 + 1.863E)^{1/2} > 15/L^2, \quad (8)$$

where R is the rigidity in GV, A is the mass number, Q the ionic charge of the particle and E its energy in GeV/nucleon.

Putting in some numbers, for instance, the geomagnetic cut-off momenta for $Z=1$ particles at different latitudes are

		Kinetic energy	
$Z = 1$	$\lambda = 0^\circ$	$cp \geq 14.9$ GeV	14.0 GeV
	$\lambda = 40^\circ$	$cp \geq 5.1$ GeV	4.3 GeV
	$\lambda = 60^\circ$	$cp \geq 0.93$ GeV	0.48 GeV

Both positive and negative particles show the same latitude effect. The cut-off energy, however, is different for positive and negative particles arriving in *non vertical* directions. Lower energy particles of negative charge would penetrate if they arrived from West, but for positive particles arriving from East a higher energy would be needed. For negative particles, the effect would be reversed. More low-energy particles than high-energy ones are found in cosmic rays, and thus this *East-West effect* predicts that, if primary CR are mainly positively charged, more CR should be detected by a Geiger telescope directed to the West than by one directed to the East. Confirmation of the East-West effect came in 1938: the majority of cosmic rays are *positively* charged.

Let us consider a detector located at the position \mathbf{x} that measures a particle of electric charge Z and momentum \mathbf{p} . To a very good approximation the past trajectory of the detected particle can be determined integrating the classical equations of motion for a charged particle in an electromagnetic field in the region around the detector (and the Earth). Reconstructing this past trajectory there are three possible results ⁴):

- (a) the trajectory originates from the Earth's surface (or deep in the atmosphere);

- (b) the trajectory remains confined in the volume $R_E < r < 1$ without ever reaching infinity;
- (c) the particle in the past was at very large distances from the Earth.

Trajectories belonging to the classes (a) and (b) are considered as "forbidden", because no primary cosmic ray particle can reach the Earth from a large distance traveling along one of these trajectories. All other trajectories are allowed. If we consider a fixed detection position \mathbf{x} and a fixed direction \mathbf{n} , to a reasonably good approximation the trajectories of all positively charged particles with rigidity larger (smaller) than a cut-off are allowed (forbidden). This is exactly true for a dipolar field that fills the entire space. In this case the solution can be written down as an analytic expression. In the more general case it is necessary to study the problem numerically. The effect of the geomagnetic field on an isotropic interplanetary flux is simply to *remove* the particles from the forbidden trajectories, without deforming the shape of the spectrum. This can be deduced from the Liouville theorem, with the assumption that the field is static.

2.2 Trapped particles

No attention was given to the "forbidden" orbits, though mathematically known, until they were truly discovered.

The discovery of these particles happened in 1958 with the Explorer 1 satellite. Being interested in cosmic rays, J. A. Van Allen's group had placed a Geiger counter on the satellite; the strange behaviour of the detected fluxes put in evidence the possibility of having trapped particles inside the magnetosphere. At the end of 1958 other measurements (Pioneer-3) gave rise to the idea that there are *two distinct belts* of trapped particles, the 'inner' and the 'outer' Radiation Belts (see figure 2).

Trapping mechanism

The trapping of such particles arises from the interaction between a moving electric charge and a static magnetic field. The principles are best expressed

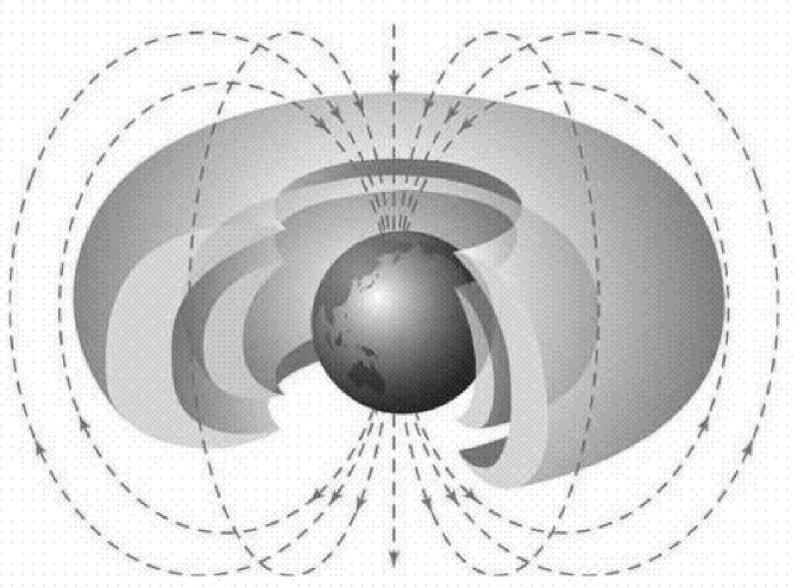


Figure 2: *Sketch of the two Van Allen belts, and the field-lines of the Earth's magnetic field.*

in terms of *adiabatic invariants*, which are quantities that do not change provided some other quantities, specifically the magnetic flux density, are steady or change sufficiently slowly.

A charged particle of mass m and charge e gyrates in a magnetic field of flux density \mathbf{B} at a *gyrofrequency* $\omega_B = eB/m$. The radius of gyration is $r_B = mv_{\perp}/Be$, where v_{\perp} is the particle velocity normal to the magnetic field.

If a particle in the magnetosphere has velocity v_{\perp} normal to the magnetic field and v_{\parallel} along it, the trajectory is a *spiral*. Provided no work is done on or by the particle, the magnetic flux through the orbit Φ_m is constant. Hence

$$\Phi_m = B\pi r_B^2 = 2\pi m E_{\perp}/e^2 B = \text{constant}, \quad (9)$$

where E_{\perp} is the kinetic energy associated with the transverse velocity component, that is

$$E_{\perp} = mv_{\perp}^2/2. \quad (10)$$

Hence, E_{\perp}/B is constant. But this is the magnetic moment of the current loop

represented by the gyrating particle

$$\mu = \text{current} \times \text{area of loop} = \frac{ev_{\perp}}{2\pi r_B} \pi r_B^2 = E_{\perp}/B . \quad (11)$$

This gives us the first adiabatic invariant of charged particles motion in a magnetic field: *the magnetic moment is constant*. It holds provided the magnetic field does not change significantly during one gyration period.

If the total velocity of the particle is v , then

$$v_{\perp} = v \sin \alpha ,$$

where α is the pitch-angle. Hence, since the total kinetic energy E is constant

$$E_{\perp}/B = E \sin^2 \alpha / B = \text{constant} .$$

Thus

$$\sin^2 \alpha \propto B .$$

As a particle moves from the equator towards higher latitude in a dipole field, it encounters increasing B and therefore the pitch-angle increases. Eventually, provided the atmosphere is not encountered first, $\alpha = 90^\circ$. Here the forward motion stops and the particle is reflected back along the field towards the equator. The point of reflection is called the *mirror point*. The total energy of the particle does not change during its motion because no acceleration mechanisms are at work; also, $E = E_{\perp} + E_{\parallel}$. Hence, for the parallel velocity

$$v_{\parallel}^2 = v^2 - v_{\perp}^2 = v^2 \cos^2 \alpha .$$

The parallel energy falls to zero at the mirror point, and the changing pitch-angle represents an alternating of kinetic energy between parallel and perpendicular components, the total remaining constant. It is the conservation of the first invariant within a non-uniform magnetic field approximating to a "magnetic bottle" that makes possible the trapping of energetic charged particles.

For a given particle, the position of the mirror points is determined by the pitch-angle as the particle crossed the equator, since

$$B_0/B_M = \sin^2 \alpha_0 / 1.0 ,$$

B_0 and B_M being the flux densities at the equator and at the mirror point respectively. The situation is sketched in figure 3.

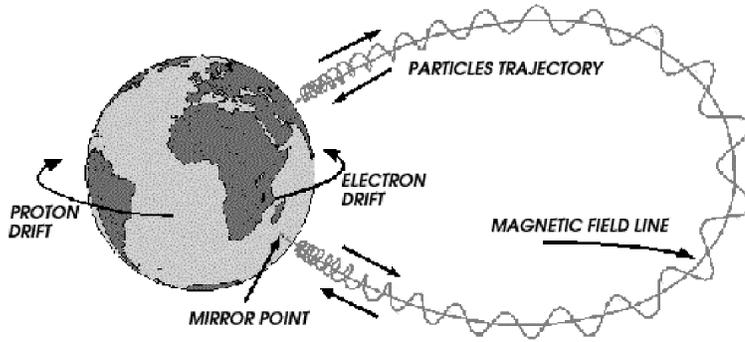


Figure 3: *Trajectories of particles trapped on closed field-lines.*

Particles will be lost if they encounter the atmosphere before the mirror point. The equatorial pitch-angles of particles that will be lost into the atmosphere at the next bounce define the *loss cone*, which will be seen as a depletion within the pitch-angle distribution.

In addition to the gyration about the field-line and the oscillation between hemispheres, a trapped particle also *drifts* around the Earth. This is due to two mechanisms; the first is that the particle finds itself in a slightly weaker field when on that part of the gyration furthest from the Earth. The ratio of the gyration therefore changes and this introduces a lateral shift of the orbit. The second cause is the curvature of the field-line, which produces a centrifugal force to which the particle responds by drifting sideways. The combination of these two effects is called the *gradient-curvature drift*.

Gradient-curvature drift sends electrons to the East and protons to the West. In a dipole-field, it would move the particles around at the same distance from the Earth and would serve merely to distribute the particles to all longitudes. However in the actual field, which is not dipolar, the drift paths are not so obvious. The theory of trapped particles shows that the path may be determined from the *second adiabatic invariant*, which states that *the integral of the parallel momentum over one bounce between mirror points is constant*. This is a property of the field configuration and also of the mirror points of the

particle and defines the surface, or *shell*, on which the particle remains as it drifts around the Earth. It holds provided the field does not change appreciably during one bounce period. Taken together, the first and second invariants define the locus of the mirror points of a bouncing and drifting trapped particle.

An important development from the second invariant is the *L parameter*, already introduced in section 2. It has the dimensions of length, taking one Earth radius as the unit. It is analogous to the distance to the equatorial crossing of a field-line (R_0 in equation 6), to which it reduces in a dipole field. Along most closed geomagnetic field lines L is constant to about 1% and thus it is useful to identify field-lines even though they be not strictly dipolar. By analogy with a dipole field, an *invariant latitude* may be defined in terms of L :

$$\Lambda = \cos^{-1}(1/L)^{\frac{1}{2}} . \quad (12)$$

This is analogous to equation 7, and through it L is commonly used to label field-lines even at high latitude where L strictly has no meaning because particles cannot be trapped on field-lines that are not closed.

The third adiabatic invariant says that *the total geomagnetic flux enclosed by the drift orbit is constant*, and this enables the effect of very slow changes of magnetospheric structure on particle orbits to be calculated. This invariant is violated if changes occur in a time less than that taken for the particle to encircle the Earth, violation quite frequent particularly during magnetic storms.

In a steady state, if no other processes were operating, particles trapped would remain trapped forever. But then one would have to ask how they became trapped in the first place. And if there is indeed a mechanism for trapping charged particles, there must obviously be a loss process too, otherwise the trapped flux would be increasingly indefinitely, which is not the case. Let us resume the present knowledge about the Inner and the Outer Radiation Belts.

Inner Radiation Belt

The source must either create particles *in situ* or introduce them by violating an invariant. In the so called *inner zone*, within about $2.5 R_E$, the

particles are created *in situ* by the decay of neutrons which themselves come from primary cosmic ray protons of very high energy:

$$5 \text{ BeV proton} \rightarrow \text{approx. } 7 \text{ neutrons}$$

$$\text{neutron} \rightarrow \text{proton} + \text{electron} + \text{antineutrino}$$

Calculations show this to be a sufficient source for the inner zone.

The loss process involves some kind of interaction with the atmosphere near the mirror points. The main inner zone loss processes are: retardation and charge exchange, particularly for the less energetic (< 100 keV) protons; nuclear collisions, particularly for protons > 75 MeV; scattering into the loss cone, particularly for electrons.

As final result of these processes, the Inner Radiation Belt is rich of protons more than electrons.

Outer Radiation Belt

The question of how the particles of the *outer zone* ($L > 3$) are created does not have such a clearcut answer. Some of the more obvious possibilities do not stand up to detailed scrutiny. The charged particles almost certainly come from outer the magnetosphere, and the main evidence for this is the increase of flux at times of enhanced geophysical activity. The particles probably enter the outer magnetosphere at the sunward cusps. Some will already be of high energy, but local acceleration in the magnetosphere is also required and it is not entirely clear how it happens.

Diffusion of pitch-angle is the cause of particle loss from the outer zone. This involves the violation of the second invariant by varying fields or interactions with electromagnetic or electrostatic waves that act in times shorter than a bounce period. When particles diffuse into the loss cone they enter the atmosphere at the next bounce and are lost there. Protons ($1.0 < E_p < 5.0$ MeV) are available in this region, though not to the same extent as the electrons ($E_e > 1.6$ MeV).

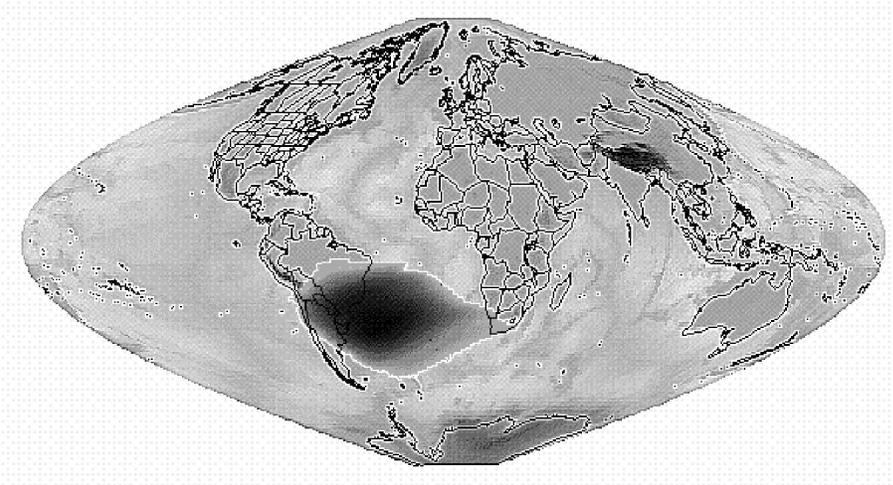


Figure 4: *The South Atlantic Anomaly SAA, visible over Brazil.*

2.2.1 The South Atlantic Anomaly

Above South America, about 200 - 300 kilometers off the coast of Brazil, and extending over much of South America, the nearby portion of the Inner Radiation Belt forms what is called the South Atlantic Anomaly (figure 4). This is an area of enhanced radiation caused by the offset and tilt of the geomagnetic axis with respect to the Earth's rotation axis, which brings part of the radiation belt to lower altitudes (below 500 km).

Satellites and other spacecraft passing through this region of space are bombarded by protons exceeding 1 GeV at a rate of 3000 particles/cm²s. This can produce problems with the operation of on-board electronic systems, and premature aging of computer, detector and other spacecraft components.

2.3 Albedo particles

Besides particles permanently trapped in stable radiation belts, the Earth's magnetic field holds inside also *albedo* particles, corresponding to the trajectories (a) of section 2.1.

Albedo particles are produced by cosmic ray interactions in atmosphere (40 km). When secondary production occurs, a fraction of particles can travel backward along the field line connected to the point of observation, and can

be rebound to space. The rigidity of these particles can be less than the local geomagnetic cut-off. Depending of their pitch-angle such particles can make only one bounce (albedo) or more than one bounce (quasi trapped). There is also some fraction of secondary particles that can be trapped some years. The fluxes of these three components are approximately constant over a wide energy range.

There are three main differences between albedo and trapped particles:

- the origin traces back into atmosphere or ground level for albedo particles, whereas trapped particles are generated at a radius corresponding to the radiation belts;
- albedo particles have a shorter flight time;
- albedo particles can have energy up to GeV, whereas the energy of trapped particles lies mainly in the MeV region.

2.4 The Earth's magnetosphere

The dipolar structure of the Earth's magnetic field is highly deformed by the influence of the Solar Wind (treated in the next section).

The Solar Wind is highly supersonic when it encounters the Earth and its magnetic field. To a rough approximation, the Earth and its associated magnetic field act as a spherical obstacle in the outflowing Solar Wind.

There is a bow shock at a distance of about $14 R_E$ from the centre of the Earth in the direction of incidence of the Solar Wind. Closer to the Earth, there is a boundary known as the *magnetopause* at a distance of about $11 R_E$ which acts as the surface of the region within which the Earth's magnetic field is dynamically dominant. This region is known as the *magnetosphere*. The Solar Wind plasma flows past the Earth between the shock wave and the magnetopause.

The most significant distortion to the Earth's dipole magnetic field (figure 5) is the fact that the magnetic field lines of the down-stream side of the Earth are stretched out by the drag of the Solar Wind. The magnetospheric cavity is stretched out into a long cylindrical region which has radius about $25 R_E$ at a distance of about $60 R_E$. This region is known as the *magnetotail*. The magnetic field lines are oppositely directed on either side of the equatorial plane, those in the northern regions heading towards the Earth, those in the southern

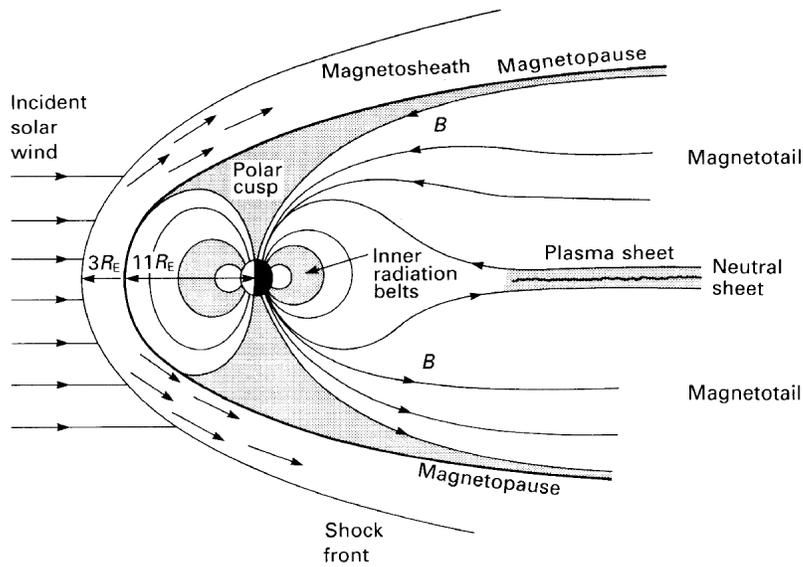


Figure 5: A schematic diagram showing the structure of the Earth's magnetosphere and the effect of the Solar Wind.

region pointing away from the Earth. Between the two regions is a thick layer of hot plasma which is known as the *plasma sheet*. Because the magnetic field lines run in opposite directions on either side of the plasma sheet, there must be a surface of zero magnetic field separating the two regions, which is known as the *neutral sheet*. Because the magnetic field changes sign through the neutral sheet, an induced electric current flows in the plasma sheet and hence particles can be accelerated in the vicinity of the neutral sheet.

This picture of the Earth's magnetosphere (again figure 5) suggests an explanation of the phenomena of the *aurorae* observed at high geomagnetic latitudes. Particles accelerated in the region of the magnetotail can drift along the magnetic field lines to high geomagnetic latitudes. Electrons with energies 0.5-20 keV entering the upper layer of the atmosphere at about 90-130 km can excite oxygen atoms producing the green 558 nm and red 630 nm lines of oxygen, characteristic of aurorae.

3 The influence of the Sun

In this section it is described the action of the Solar Wind to the flux of primary cosmic rays, and it is shown that also the Sun is a source of particle radiation.

The Sun's energy, originated by fusion reactions inside, radiates in all directions, maintaining a steady level (quiet Sun). The photosphere, visible surface of the Sun, has a temperature of $T = 6000 \text{ }^\circ\text{K}$, but the overlying corona has a T exceeding $10^6 \text{ }^\circ\text{K}$. At these temperatures, part of the ionized gas of the solar ambient has speed enough to escape the solar gravitational attraction, flowing out through the solar system as a Solar Wind.

3.1 The Solar Wind

The Solar Wind is now easy to detect with particle detectors carried on satellites, but before the space age the indications of its presence were less direct. The existence of the Solar Wind was firmly established only in 1960, by summing up several evidences:

Sunspots: after Galileo's discovery of the telescope in 1609, it became clear that the sunspots were truly attached to the Sun because they moved across the Sun's face according to a 27-day rotation period. The number of sunspots visible at any time is highly variable. It was found that they occur in cycles, their number increasing or decreasing over intervals of about 11 years. The Zurich sunspot number, still our most direct indicator of the solar activity, had a periodicity similar to that of the geomagnetic disturbances, but the physical link could not be easily found.

Magnetic storms: disturbances in Earth electrical power systems and telecommunications, often accompanied by aurorae in polar regions. They were correlated with the 27-day periodicity of the sunspot number, but for many years this link was thought accidental. The correlation, however, persisted for more than 400 cycles.

Antimodulation of CR: neutron monitors at ground found an anti-correlation between the particle fluxes and the sunspot number (see figure 6). Over the 11-year solar cycle, when sunspot number increased, the CR intensity decreased. The conclusion was that the solar activity, rather than

producing more cosmic rays, was preventing galactic cosmic rays from reaching the Earth.

The last piece of the puzzle came from observations of comets.

Comets: a mixture of frozen water, frozen gases and dust. Comets generally have two tails. Type I tails are composed of gaseous ions such as CO^+ and OH^+ ; type II tails, in contrast, contain many dust particles that were originally frozen into the icy comet but were then released as the binding ice evaporated. Under the pressure of the solar radiation, the dust particles get blown away from the comet and become visible by the sunlight they reflect. The ions are much smaller than the dust particles and are less affected by solar radiation.

Type I tails were studied in detail. The movement of the dense clumps of ionized gas away from the head of the comet were understood in terms of collisions with particles emitted continuously from the Sun. The Solar Wind and the interplanetary magnetic field force the ionized gas to stream behind the comet. Following the movements of the tails it was possible to infer speed and number of particles in the Solar Wind.

The Solar Wind is composed of protons and electrons, in a neutral stream. There is also a component of He and heavier elements, in less amount. Its high temperature ensures that the gas is highly ionized. Within the stream is a magnetic field, so there is a magnetically complex region where the solar region flows by the Earth and the Earth's own magnetic field, as we have seen already. Because of the Sun rotation, the Solar Wind does not simply stream out along straight radial directions, but follows curved spiraling paths in what it is called the 'garden-hose' effect. The Earth thus encounters the Solar Wind coming from a direction slightly ahead of the direct Earth-Sun line. The Solar Wind has been detected as far out as 40 AU from the Sun. At 50-100 AU it is thought that the Solar Wind terminates abruptly in a "shock", a complex boundary between the interplanetary and interstellar regions.

3.2 Active Sun

The Solar Wind is a manifestation of the quiet Sun. Solar Cosmic Rays (Solar Energetic Particles) are instead a short-lived manifestation of the active Sun, and are associated to energetic solar events.

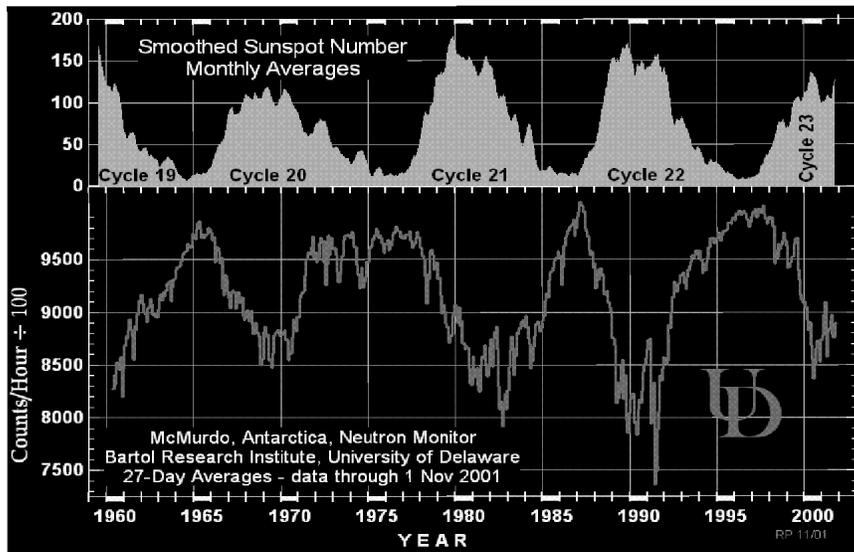


Figure 6: *The correlation between the sunspot number and the intensity of neutrons measured by the Antarctic McMurdo neutron monitor.*

The two groups of particles are distinguished by their energy: the Solar Wind protons have energy in the keV region, whereas Solar CR protons span the MeV energy interval.

Solar Energetic Particles come from two different sources, often occurring together:

- Solar Flares: until the 90ies thought to be responsible of the most intense SEPs and geomagnetic storms. The Solar Flare is an explosive release of energy (both electromagnetic and charged particles) within a relatively small (but greater than Earth-sized) region of the solar atmosphere;
- Coronal Mass Ejections (CMEs): violent eruptions of coronal mass, known to be the very responsible of particle acceleration. Often, not always, associated to a flare. The fast CME explosion in the slow Solar Wind produces a shock wave, which accelerates particles.

4 What else arrives to Earth?

Among the many surprising and important discoveries derived from experiments on spacecraft over the past forty years, there is the presence of an energetic charged particle component in the heliosphere, not of solar or galactic origin, at energies around few tens of MeV/nucleon. Its presence in the heliosphere is an independent probe for both interplanetary electrodynamic investigations - especially solar modulation - and probably the most direct means for determining the elemental and isotopic composition of those neutral atoms in the local interstellar medium that have high first ionization potentials (He, N, O, Ne, Ar, ...). Throughout the literature many authors have called the anomalous component discussed herein the “Anomalous Cosmic Rays (ACR)” which is finally an unfortunate misnomer since the radiation is not cosmic rays, as it will be explained later on.

In 1972, as the galactic cosmic ray intensity was rapidly recovering from intense modulation associated with the 11-year solar cycle, it was reported ⁵⁾ that the helium differential spectrum below ~ 60 MeV/nucleon could not be accommodated within the current solar modulation theory. Indeed, below ~ 60 MeV/nucleon the observed quiet time helium flux at 1 AU far exceeded the modulated proton flux of galactic origin, whereas it was well established that the He/H ratio for galactic components was ~ 0.2 . In a series of observations from 1973 till 1975 ⁶⁾ it became clear that this excess helium was pure ^4He and therefore was a local component different from the galactic cosmic radiation which displayed at low energy an E^{+1} spectrum.

In 1973 was then discovered ⁷⁾ a flux of oxygen which exceeded the expected low energy cosmic ray oxygen flux by a factor ~ 40 , and again it was reported ⁸⁾ that nitrogen and oxygen at low energies had anomalously high fluxes while carbon showed the expected modulated spectrum.

4.1 Mechanism for Anomalous Cosmic Rays

Physicists ⁹⁾ recognized that these elements with anomalously high observed abundances were characterized by a high first ionization potential (FIP) and proposed a model wherein neutral interstellar atoms enter the heliosphere with a relative speed of $\sim 25 \text{ km s}^{-1}$. Those with highest ionization potential penetrate deep into the heliosphere before undergoing either photo-ionization or

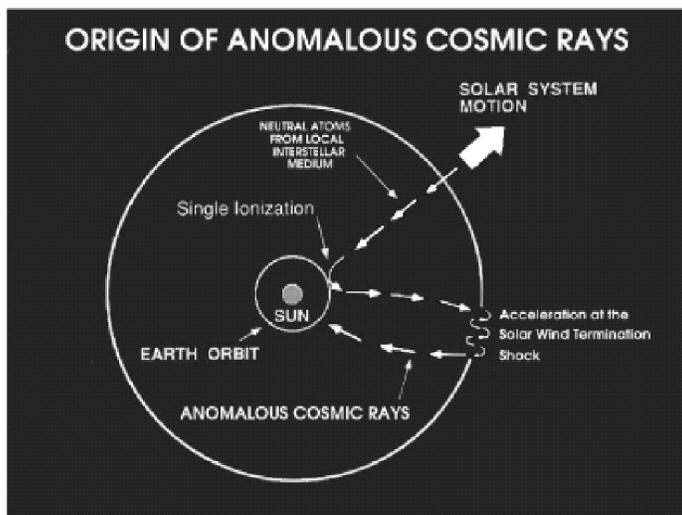


Figure 7: *Sketch illustrating the model proposed by Fisk to account for the anomalous nuclear component.*

charge exchange to become singly-charged atoms. As singly charged atoms they are then picked up in the Solar Wind and carried to the outer heliosphere where they undergo an unspecified local acceleration. A portion of these accelerated nuclei then return to the inner Solar System as modulated pseudo “cosmic rays” which, since are only singly charged, have much higher magnetic rigidities than the fully stripped galactic cosmic rays of the same energy and can therefore arrive closer to the Earth. Figure 7 is a sketch of this scenario.

Following the discoveries of ^4He , N and O, other high FIP atoms with anomalously high abundances have been found, whose physical properties fall within the proposed model (known as the Fisk model).

The most likely mechanism for the acceleration of the anomalous component is, at the moment, a heliospheric termination shock. After the acceleration, the anomalous component propagates inward from the distant heliosphere and undergoes diffusion, drift, convection and adiabatic deceleration along with the galactic cosmic rays over the 11-year solar modulation cycles. Since the low energy differential spectra of this component have the approximate form $\propto E^{-\gamma}$, where γ is the spectral slope, the fluxes of this component observed at 1

AU are extremely sensitive to the solar modulation, changing by more than two orders of magnitude over the 11-year or 22-year solar modulation cycle. More recent results confirm a modulation of AC ions not only with the intensity of the solar activity (11 years) but also with the polarity of the solar magnetic field (22 years) ¹⁰⁾.

The Fisk model makes an important prediction about the anomalous component, and it is that AC particles should be *singly ionized*. There have been numerous attempts to infer the charge state of the AC from studies of its response to solar modulation ¹¹⁾; all these attempts have favored a singly charged AC, nevertheless such arguments are obviously unsatisfactory in some respects, especially because the conventional solar modulation theory provide only an incomplete and approximate account of the dynamic heliospheric processes which affect cosmic rays.

After these discoveries, later observations showed an evidence of *multiply charged* oxygen in the anomalous radiation ¹²⁾. This new effect was explained again in the frame of the Fisk model: multiply charged oxygen exists as a result of additional electron stripping during the acceleration process at the Solar Wind termination shock, after the initial ionization, pick-up, and transport by the Solar Wind. This placed new constraints on the acceleration time-scale and so helped to explain how AC atoms can be accelerated to high energies. Detailed modeling of the acceleration and stripping processes are now existing ¹³⁾.

More recently it has been shown that further electron stripping also occurs during the acceleration of other AC ions. In particular, multiply charged *Ne* and *N* have been observed ^{14, 15)}. At low energies, AC ions tend to be more singly charged, while at higher energies ($E \geq 20$ MeV/nucleon) they are predominantly multiply ionized. This energy dependence can provide a strong constraint on acceleration models.

4.1.1 *The trapped anomalous component*

The discovery of the anomalous nuclear component in 1972 did stimulate theoretical interest. In 1977 it was ¹⁶⁾ first outlined the mechanism by which this anomalous component could become *trapped* in the inner magnetosphere: the high mass-to-charge ratio of singly ionized AC nuclei enables them to penetrate deeply into the magnetosphere. The grammage required to strip one such nu-

cleus at ~ 10 MeV/nucleon is much smaller than its range, so AC nuclei whose trajectory takes them near to a low altitude mirror point easily encounter sufficient grammage to be stripped of remaining orbital electrons. After stripping, the particle gyroradius is reduced by a factor $1/Z$ (Z =atomic number) and the ion can become stably-trapped. Its lifetime is then determined only by energy loss at subsequent encounters with the residual atmosphere at the low altitude mirror points. Moreover, since its lifetime typically corresponds to hundreds of drift periods around the Earth, the flux of trapped particles greatly exceeds the AC flux outside the magnetosphere.

The SAMPEX spacecraft ^{17, 18)} has provided the first detailed look at trapped AC with high resolution electronic detectors, and has found a band extending from the tip of South America to the southern tip of Africa (SAA) which consists of anomalous cosmic rays that have become trapped in the Earth's magnetic field after losing some or all of their remaining electrons.

Trapped ACRs form a specific radiation belt. This radiation belt includes significant abundances of N and Ne, but very little C or other elements. The L – *shell* distribution of the observed O and N is sharply peaked at $L \sim 2$ (see figure 8, in a very narrow region (FWHM in $L \sim 0.3$).

The composition of the trapped ions has been a key point to identify their origin in the SAMPEX data. The trapped composition was in fact compared with the compositions of several possible source populations at $L \sim 2$. They found that the trapped composition was generally consistent with that observed for the anomalous cosmic component in the interplanetary space, but not with other possible sources of magnetospheric particles.

Recent measurements of the trapped AC component during a whole solar cycle have shown the presence of a strong solar modulation, in accordance to the corresponding untrapped component ¹⁹⁾.

5 Summary of cosmic rays

The previous sections have shown that there are three different types of cosmic rays ²⁰⁾:

- (a) **Galactic Cosmic Rays (GCR)** originating far outside our solar system. The source of the very energetic (GeV range) galactic cosmic rays is not known. Occasionally (one event per square kilometer per century) GCR

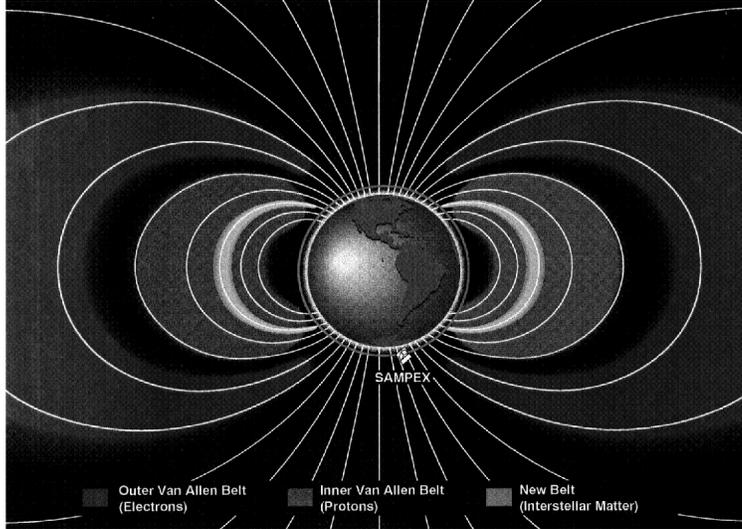


Figure 8: *Trapped nucleon belt at $L \sim 2$ from anomalous nuclear component.*

carry energies as high as 10^{20} eV.

GCRs are the most typical cosmic rays, and their flux in the solar system is modulated by the solar activity: enhanced Solar Wind shields the system from these particles. The effect is clearly seen from the cosmic ray measurements covering several solar cycles. Quite interesting fact is that also the 22-year modulation is seen in the recovery of GCR intensities: during even cycles it is completed much faster than during odd cycles.

- (b) **Solar Cosmic Rays (SCR)**, or Solar Energetic Particles, originating mostly from solar flares. Coronal Mass Ejections and shocks in the interplanetary medium can also produce energetic particles. SCR particles have energy of up to several hundreds of MeV/nucleon (sometimes up to few GeV/nucleon). The composition is similar to Galactic Cosmic Rays: mostly protons, about 10% of He and $< 1\%$ heavier elements. During strong solar flares (if optimally located on the Sun), the flux of CR at the Earth orbit can increase for some hundreds percent during hours/days, leading to a SEP event.

- (c) **Anomalous Cosmic Rays (ACR)**, originating from the interstellar space beyond the heliopause. The Anomalous Cosmic Rays are produced by neutral atoms in the interstellar space, which leak into the heliosphere, get ionized by solar UV radiation or charge exchange with the Solar Wind, are picked up by the Solar Wind and convected back to the outer heliosphere, are accelerated, e.g., by the Solar Wind termination shock, and diffuse and drift into the inner heliosphere as cosmic rays.

Anomalous Cosmic Rays differ from others by their composition (20). While in GCRs and SEPs there are much more protons than helium, and equal amounts of oxygen and carbon, in ACRs there are more heliums than protons, and much more oxygen than carbon. This can be explained with the production scheme outlined in figure 7, which selects only those elements that are predominantly neutral in the interstellar medium: it is known that the first ionization potential of such atoms as oxygen and helium is high. Another difference is that ACRs are singly charged, while GCRs have been stripped of their electrons during their millions year passage through the Galaxy. Also, the energy of ACRs does not reach the highest values found in GCRs.

The acceleration of cosmic rays to the observed energies is still partly an open question, especially at the ultra high energy range. Two mechanisms that surely play a role are the Fermi acceleration (suggested first by Enrico Fermi, 1949) and magnetic pumping (suggested by Hans Alfvén, 1963).

The Earth's atmosphere protects us from being exposed to these particles. As a cosmic ray enters the atmosphere, it will collide with a particle in there (usually a nitrogen or oxygen molecule), or interact with the molecules, exciting them and thus causing an outer space influence upon the Earth's environment.

Besides these three families, the Earth magnetic field holds inside particles temporary and stably trapped.

The trapping regions of high-energy charged particles surrounding the Earth are called Radiation or Van Allen Belts. The Inner one, located between about $1 R_E$ and $3 R_E$ in the equatorial plane, contains primarily protons with energies exceeding 10 MeV. This is a fairly stable population but it is subject to occasional perturbations due to geomagnetic storms, and it varies with the

11-year solar cycle. The source of protons in this region is the decay of cosmic ray induced albedo from the atmosphere.

As a result of the offset between the Earth's geographical and magnetic axes, the Inner Belt reaches a minimum altitude of about 250 km above the Atlantic Ocean off the Brazilian Coast. This South Atlantic Anomaly occupies a region through which low-orbiting satellite frequently pass. Energetic particles in this region can be a source of problems for the satellites and astronauts.

The Outer Belt contains mainly electrons with energies up to 10 MeV. It is produced by injection and energization events following geomagnetic storms, which makes it much more dynamic than the Inner Belt (it is also subject to day-night variations). It has an equatorial distance of about 3 - 9 R_E . The 'horns' of the outer belt dip sharply in towards the polar caps.

Recently a new belt has been found within the Inner Belt. It contains heavy nuclei (mainly oxygen, but also nitrogen and helium, and very little carbon) with energies below 50 MeV/nucleon. The source of these particles is Anomalous Cosmic Rays of interstellar origin.

6 Cosmic rays missions

Direct cosmic rays detection is nowadays performed with three basic technologies: balloon-borne and satellite-borne detectors, and instrument placed aboard space stations. For all of them, the knowledge of the radiation environment related to the mission is of vital importance.

For balloon missions the choice of the geographic location is essential for the physics that the mission intends to study. As it was shown in this article, indeed, every location on the globe has associated a geomagnetic cut-off rigidity below which no vertically arriving particles can penetrate the Earth's magnetic field. In addition, background calculations due to albedo or trapped particles, whose fluxes must be evaluated with good precision, represent an important step of the data analysis.

For satellite missions the choice of the orbit is a basic parameter for the scientific objectives, and care should be taken in choosing the one that corresponds to the optimal radiation conditions for the physics studied. Polar orbits allow charged low energy cosmic rays to be detected at high latitudes, whereas equatorial orbits are screened from this radiation. Detectors in space studying the gamma radiation would more realistically need equatorial orbits in

order to be screened from cosmic rays, while a different choice should be made for a mission studying the charged component of the cosmic radiation. Also the knowledge of the background conditions is important in order to produce reliable results.

All such considerations hold also for space station missions, where anyway one more important item should be taken seriously into account: the safety of humans aboard. Manned mission last several years, and astronauts are exposed to high levels of dose while living in the station. A passage in the radiation belts is a heavy hazard for the human safety and, in addition, the radiation environment changes at times abruptly due to the solar activity, and countermeasures should be taken to guarantee the protection of humans aboard.

7 Conclusions

Cosmic rays travel much and are distorted before reaching the Earth. The Earth magnetic field, the Sun and the atmosphere influence strongly the primary cosmic ray flux, and create a region of space around the Earth where different families of particles exist.

The knowledge of the radiation environment related to a space mission is necessary when choosing the strategy of the mission from the scientific point of view, and it is important for the data analysis and the interpretation of the results.

References

1. R. Sparvoli, NINA: a New Instrument for Nuclear Analysis of primary cosmic rays, PhD Thesis (1998).
2. M. W. Friedlander, Cosmic Rays, Harvard University Press (1989).
3. J. K. Hargreaves, The solar-terrestrial environment, Cambridge University Press (1995).
4. P. Lipari, astro-ph/0101559 (2001).
5. M. Garcia-Munoz, G. M. Mason, J. A. Simpson, ApJ **182**, L81 (1973).
6. M. Garcia-Munoz, G. M. Mason, J. A. Simpson, ApJ **202**, 265 (1975).

7. D. Hovestadt *et al.*, Phys. Rev. Lett. **31**, L650 (1973).
8. F. B. McDonald *et al.*, ApJ **187**, L105 (1974).
9. L. A. Fisk, B. Kozlovsky, R. Ramaty, ApJ **190**, L35 (1974).
10. P. Mäkelä, *et al.*, Proceedings XXV ICRC Durban **2**, 281 (1996).
11. T. T. von Rosenvinge, F. B. McDonald, Proceedings XIV ICRC **2**, 792 1975.
12. R. A. Mewaldt *et al.*, ApJ **466**, L43 (1996).
13. J. R. Jokipii, ApJ **466**, L47 (1996).
14. R. S. Selesnik *et al.*, Proceedings XXV ICRC **2**, 269 (1996).
15. B. Klecker *et al.*, Proceedings XXV ICRC **2**, 273 (1996).
16. J. B. Blake, L. M. Friesen, Proceedings XV ICRC **2**, 341 (1977).
17. D. N. Baker *et al.*, IEEE Trans. on Geoscience and Remote Sensing **31-3**, 531 (1993).
18. J. R. Cummings *et al.*, Geophys. Res. Lett. **20**, 2003, (1993).
19. R. S. Selesnik *et al.*, Proceedings XXV ICRC **2**, 305 (1996).
20. R. A. Mewaldt, A. C. Cummings, and E. C. Stone, EOS **75**, N. 16 (1994).

**The antimatter component of cosmic rays and the PAMELA
experiment**

Piero Spillantini
INFN & University, Firenze, Italy

ABSTRACT

In this lesson it will be adopted the following scheme:

- (a) - The antimatter component of cosmic rays:
 - (a.1) Predictions of the theories
 - (a.2) Status of the experimental observations
 - (a.3) Observation programs for the next future
- (b) - The PAMELA experiment

1 The antimatter component of cosmic rays.

1.1 Prediction of the theories.

I will begin my lesson following an historical approach.

The chronicle of the antimatter search can be summarized according the following table:

- 1928 prediction of antielectron (Dirac)
- 1932 discovery of positron in cosmic rays (Anderson)
- 1954 "antiproton induced" events in cosmic rays (Amaldi)
- 1955 antiproton manufactured in laboratory (Chamberlain et al.)
- 1965 antideuteron manufactured in laboratory (Massam et al.)
- 1960's Baryon Symmetric Cosmologies (Klein, Alfven,..)
- 1967 Sakharov's conditions (Sakharov)
- 1970's Baryon Symmetric Cosmologies (Stecker,...)
- 1970's gamma-ray "evidence"
- 1979 discovery of antiprotons in cosmic rays (Bogomolov et al., Golden et al.)
- 1996 antihydrogen manufactured in laboratory
- ???? ? antinuclei in cosmic rays? (??)

In the period between the 30's and the 60's the fundamental question that urged the cosmologists was:

how can the Universe contain equal amounts of 'particles' and 'antiparticles', as implied by the rigorous symmetry of the fundamental laws of the Nature?

The task was that of constructing mechanisms for separating particles and antiparticles on a cosmological scale. Many works were dedicated to this task, but it could not be solved at that time. The Big Bang models based on statistical fluctuations resulted in astronomical objects of very tiny mass,

less than 10^{-30} of the mass of the Galaxy. Furthermore they could not avoid the so called 'annihilation catastrophe' at the very beginning of the history of the Universe, leading to an extremely low ratio of 10^{-18} between Baryons and Photons at the present time, to be compared to the observed 10^{-9} .

The CP violation observed in the weak interactions in 1964, allowed Sakharov to formulate the hypotheses for achieving a sufficient Baryon Asymmetry in the Early Universe:

- (1) Baryon decay allowed,
- (2) CP violation allowed,
- (3) a period out of equilibrium.

Of these assumption, the first one has not still been observed, the second one occurs in kaons decay, not with the required strength, and the third one is surely satisfied in the hot Big Bang models.

These three Sakharov's conditions are still the best that we have for conceiving the possibility of an "all matter" Universe. They offers a solution to the particles-antiparticles separation problem in a Baryon Symmetric Universe, according to the following scheme:

- (1) CP is spontaneously violated,
- (2) there can be domains with either all particles or all antiparticles,
- (3) and inflation can increase these domains to the astronomical scale.

The conclusion that we can formulate now from the history of the prediction of the theories is the following:

- The theory needed to support a Baryon Asymmetric Universe is far away from being complete.
- Our present understanding does not forbid a Baryon Symmetric Universe.

Concerning this last point it must be pointed out that the observation of an "all matter" Universe is a local phenomenon. It concerns only our supercluster of galaxies, a volume of 10^{-8} of the volume of the Universe, where the ratio antibaryon/baryon is observed to be less than 10^{-5} in the hypothesis of well-mixed baryon and antibaryon gas systems.

1.2 Status of the experimental observations.

In the lack of a guiding theory, we can only investigate by direct or indirect observation for understanding if Universe is or is not symmetric.

Indirect observations can be obtained by studying the spectrum of the gamma ray radiation arriving to us from the space. An abundant presence of antiprotons in the Universe should give a 'bump' in the gamma ray spectrum, due to the γ 's coming from the decay of the π^0 's in the annihilation of the antiprotons with the interstellar and intergalactic matter. The absence of such a 'bump' allows putting limits on the fraction of existing antiprotons. These limits obviously depend from the volume of space considered around us. In the hypothesis that the gasses of protons and antiprotons are well mixed, the following figures are obtained:

- less than 10^{-15} in the particle clouds in the Galaxy;
- less than 10^{-10} in the halo of the Galaxy;
- less than 10^{-5} in our cluster of galaxies.

Until about 1995, it was claimed that the big bump observed in the high-energy diffuse cosmic X-ray background, in the energy region between 1 and 10 GeV, could be interpreted as a red-shifted signal of antiproton proton annihilation processes, with the γ 's of the π^0 decay arriving to us from cosmological distances, at red shifts of about 100. The existence of such bump was based on several balloon measurements conducted at the top of the atmosphere, very difficult and suffering of huge corrections and systematic uncertainty.

In 1995 the much more clean data of the COMPTEL experiment on board of the Compton Gamma Ray Observatory in Orbit around the Earth, made the bump disappear. With the bump disappeared also any possible indirect signal of a significant presence of antiparticles in the Universe. We can therefore only rely on direct observation of antiparticles reaching us from the space. These observations must necessarily be performed outside of the Earth atmosphere, by direct detection of the antiparticle component of cosmic rays on board of stratospheric balloons or of satellites.

This component consists of positrons, antiprotons and antinuclei.

- Positrons are not very useful for this kind of investigation. They suffer of the huge background coming from many astrophysical processes, and it is very difficult to pick up a significant contribution from extragalactic sources.
- Antiprotons are relatively abundant in cosmic rays, because they are produced as secondary in the interaction of protons with the interstellar matter. This is a background for the detection of possible extragalactic contributions. This background decreases at energies exceeding 10 GeV, and at much higher energies, around or more than 100 GeV, antiprotons could be an enough sensible probe of extragalactic contributions, either coming from 'dark matter' processes, or from the diffusion of antiprotons from 'all antiparticle' domains in the Universe. The 'dark matter' processes should show up as a 'bump' in the antiproton energy spectrum, while the antiproton diffusion from 'all antiparticle' domains should give a smooth rising of the antiproton/proton ratio with the energy.
- For what concerns the antinuclei, we have no idea of their abundance, neither of the possible background. If they diffuse from an 'all antiparticle' domain, we can guess that their chemical abundance could be similar to our 'all matter' domain. Therefore the antihelium nuclei should be much more abundant than the other ones, and are the first candidates we should look for. However there are two important difficulties that antinuclei should overcome to reach us:
 - ◊ They cannot travel to us following a direct path. The intergalactic magnetic fields brakes their diffusion, and the probability of arriving is proportional to their energy. Therefore those antinuclei that should be much more abundant, with energies of one or a few GeV/nucleon, could have a low probability of arriving, while those that have higher energies and could arrive, could be much less abundant at the origin.
 - ◊ The probability for an antinucleus to reach our position in the Galaxy could be also diminished by the difficulty it will find for winning the galactic wind flowing from the Galaxy. It is the analogous of the solar wind preventing the less energetic galactic cosmic rays to reach our position in the solar system. Unfortunately also the galactic wind contrasts the penetration of the potentially most abundant less energetic antinuclei.

At present all the direct measurements of the antiparticle component of cosmic rays were made by balloon borne experiments, and gave results for the positron and for the antiproton energy spectra up to about 50 GeV. The sta-

tistical and systematic errors of these measurements are somewhat large, up to about 30% for the highest points in energy. Inside these errors the fluxes of both these components can be justified by production of proton interactions on the interstellar matter (see the figures of the antiproton/proton and positron/(positron+electron) ratios in the *Balloon experiments* lesson). No antihelium nuclei have been detected, with a sensitivity down to 10^{-6} on the antihelium/helium ratio.

We can therefore conclude that at present, in the limits of the errors and of the explored energies, no experimental indications were obtained for the existence of significant quantities of cosmological antiparticles in the Universe.

1.3 Observation programs for the next future.

In order to obtain a more significant answer to the problem of the symmetry or of asymmetry of the Universe in its content of particles and antiparticles it is necessary to push the experimental investigation in the following directions:

- increase the statistics of the observation, either by long duration balloon borne experiments or by satellite borne experiments;
- the increase of the statistics will allow also to reach a higher energy in the study of the energy spectra;
- however this cannot be easily obtained by balloon borne experiments, because the secondary production on the residual air on top of the balloon prevents the possibility of exceeding 50 GeV without being flood by the background; therefore, for reaching higher energies it is mandatory to perform satellite borne experiments.

These directions are at the basis of the short coming experiments devoted to the study of the antiparticle components of cosmic rays:

- The BESS experiment will go on to increase the statistics in the measurement of the antiproton component at relatively low energies, from a few hundreds MeV up to a few GeV. At these energies the background due to the production of secondary antiprotons in the residual atmosphere is negligible. Therefore a significant increase of the statistics can be obtained by going from the present balloon borne experiment duration, of about 20 hours, to the long duration

balloon experiments in the Antarctic continent. The BESS program foresees to begin such experiments in 2003.

- The PAMELA experiment will be the first one that will study the antiparticle component in orbit around the Earth. It will be launched on board of a Russian satellite at the end of the year 2002, and will collect data for three years, allowing to measure the positron spectrum up to 270 GeV, and the antiproton spectrum up to 190 GeV. To the description of this experiment it will be devoted the last part of this lesson.

- A much larger acceptance experiment will follow up, the AMS-02 experiment on board of the International Space Station, starting in the year 2004 or 2005. Also if probably it will not allow extending very much the energy spectra for positrons and antiprotons, it will increase the sensitivity for hunting antinuclei down to about 10^{-10} in the antihelium/helium ratio. AMS-02 will be described in the lesson of Prof. Battiston.

2 The PAMELA experiment.

The PAMELA experiment is the most important activity of the international collaboration known by the name 'Wizard'. This collaboration was constituted 15 years ago for performing the WIZARD experiment at the cosmic ray facility ASTROMAG foreseen on board of the International Space Station FREEDOM. The ASTROMAG facility was based on the use of a powerful magnetic system based on superconducting coils. ASTROMAG was one of the two main programs foreseen in the general program for cosmic rays research recommended by the NASA Cosmic Ray Program Working Group established by NASA in 1984, following the recommendation made in 1982 by the National Academy of Sciences of USA. The other foreseen main program was an explorer sent in the interplanetary space for studying in detail the low energy portion of the galactic cosmic rays. This probe, known with the name ACE, is now working in space since a few years, sending the best low energy cosmic rays data until now collected. Instead the ASTROMAG facility could not be realized for the cancellation of the FREEDOM Space Station by the USA. For this facility were already selected three experiments, covering the most important open questions in the study of the galactic cosmic rays:

- The SCINATT experiment, proposed by a Japan-USA collaboration, dedicated to the study of the chemical composition of cosmic rays at energies up to 10^{16} per nucleus.
- The LISA experiment dedicated to the study of the chemical and isotopic composition of cosmic rays.
- The WIZARD experiment, based on an Italian-American collaboration, dedicated to the study of the antiparticle component in cosmic rays and to the search for antinuclei.

After the cancellation of the FREEDOM Space Station the WIZARD collaboration did not disband and decided to go on on the proposed researches by balloon borne and satellite borne experiments. The first balloon borne experiment was launched in 1989, afterwards the collaboration acquired an increasing experience, allowing to afford the much more complex PAMELA experiment (fig.1). The balloon experiments performed by the WIZARD collaboration are

PAMELA Background



Balloon exp's

MASS (89)
 MASS1(91)
 TR93 (93)
 CAPRICE (94)
 CAPRICE (97)
 CAPRICE (98)
 [CAPRICE (02)]



Low En. C.R.
 in orbit

NINA (98)
 NINA2 (00)



Life Science
 on MIR and ISS

SIL-EYE-1 (95)
 SIL-EYE-2 (97)

Figure 1: The Wizard background

shortly described in an other lesson of this course. The collaboration conducted also a number of experiments dedicated to the study of the low energy portion of the solar cosmic rays either on board of satellites (NINA and NINA2) or on the MIR Space Station (SIL-EYE-01 and SIL-EYE-02). These experiments are described in dedicated lesson of this course.

In the last part of this paragraph it will be described the PAMELA experiment. The main scientific objectives of PAMELA are:

- 1 - measurement of the energy spectrum of antiprotons up to 190 GeV and down to 80 MeV,
- 2 - measurement of the energy spectrum of positrons up to 270 GeV and down to 50 MeV,
- 3 - search for antinuclei with a sensitivity of 3×10^{-8} in the antihelium helium ratio,
- 4 - measurement of the energy spectrum of protons up to 700 GeV and down to 80 MeV,
- 5 - measurement of the energy spectrum of electrons up to 2 TeV and down to 50 MeV.

The PAMELA instrument will fly in a highly inclined (quasi polar) orbit, and will collect data during the period of transition of the solar activity from its maximum to its minimum, down to very low energies, overlapping with the energies studied with the NINA and NINA2 experiments. Therefore to the above main objectives of PAMELA the following by-product investigations can be added:

- 6 - modulation of galactic cosmic rays by the solar wind,
- 7 - study of the Solar Energetic Particle fluxes as a function of the time and of the energy,
- 8 - stationary and disturbed fluxes of particles in the magnetosphere.

The scheme of the PAMELA instrument is shown in fig.2, where also the main features of the detectors are reported. The instrument is constituted by a telescope of particle sensors, based on a magnetic spectrometer complemented by several detectors. The spectrometer is a system of five permanent magnets interleaving 6 plane of very sophisticated silicon sensors. Before the spectrometer the particle will encounter an extremely compact Transition Radiation

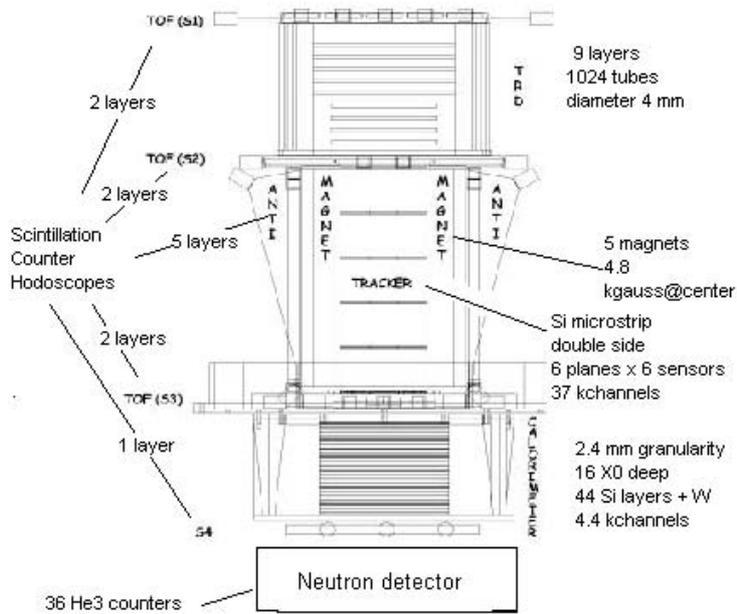


Figure 2: Scheme of the PAMELA instrument.

Detector (TRD) that will select the electromagnetic component of cosmic rays. After the spectrometer the particle will enter a very compact and deep imaging calorimeter, that will supply the detailed description of the interaction of the particle inside its volume, allowing to identify the nature of the particle. A set of scintillation counter hodoscopes (S1, S2 and S3) will supply the triggers, will measure the charge of the particle and its times of flight between the different hodoscopes. A penetration counter (S4) on the bottom of the imaging calorimeter will measure the flux of particles escaping from the calorimeter, and a neutron counter system (ND) will measure the number of neutrons es-

caping from the interactions in the calorimeter volume. Finally a system of several anticoincidence scintillation counters will protect the apparatus from external background. The total mass of the apparatus is 480 kg, its electric power consumption 345 W, its geometric factor $20.5\text{cm}^2\text{sr}$. The spectrometer has a Maximum Detectable Rigidity (MDR) exceeding 740 GV/c.

The PAMELA instrument will fly on board of the RESURS-DK1 Russian satellite in a highly inclined elliptic orbit (fig.3). Distinctive features of the

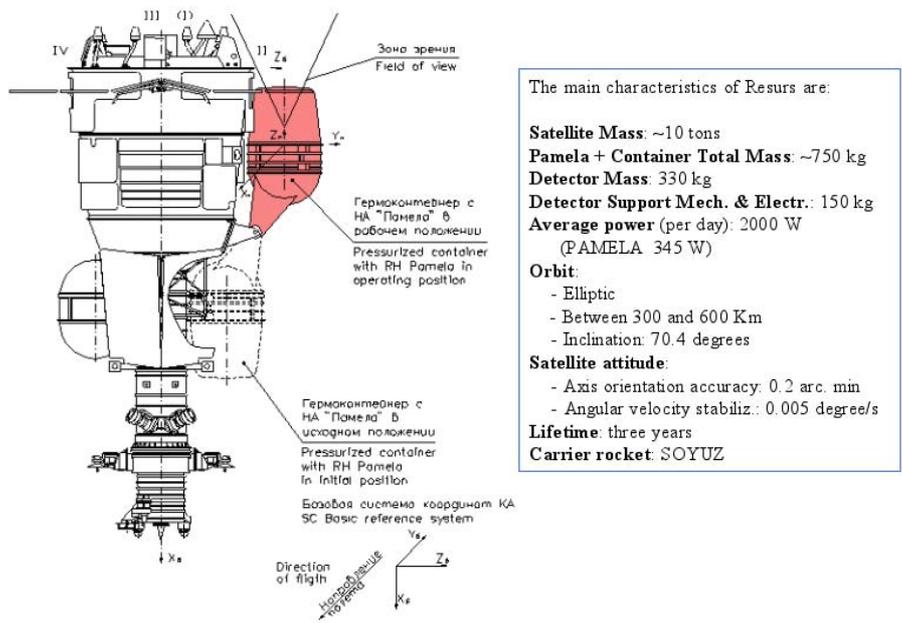


Figure 3: Scheme of the PAMELA instrument.

PAMELA instrument are the following:

- The wide energy range covered with the same instrument. This is obtained maximizing the MDR of the spectrometer (the position of the going through particle in each plane of the tracker is measured with a precision of $3\mu\text{m}$) and minimizing the thickness of the triggering scintillation counters ($\leq 0.7\text{cm}$).
- A robust separation between electromagnetic and hadronic particles, better than a part on 10^5 @ 90% efficiency. To obtain such separation the Imaging Calorimeter is 16 Xo deep, is highly granular and is complemented by the TRD,

the penetration counter S4 and the Neutron Detector.

- The Imaging Calorimeter can be calibrated in flight by selecting electrons in the TRD and measuring their momentum in the spectrometer.
- The multiple scattering in the magnetic spectrometer has been kept as low as possible (it contributes with less than 4% to the measurement error) by supporting the silicon sensors only at their edges, so that no other material than silicon is added on the particle path.
- The time of flight of the particle through the telescope is measured several times in order to improve its precision, but also for rejecting background.

Several prototypes of the various detectors have been constructed and tested on particle beams before affording the final construction. The performance of the final prototypes is:

♣ Magnetic spectrometer: the magnetic field supplied by the five magnets is 4.8 kgauss, uniform in all the magnetic volume of the spectrometer ($16 \times 14 \times 45 \text{cm}^3$). The tracker gives the position of the particle in each plane on the bending view with a standard error of $3\mu\text{m}$. [However, the above quoted value of 740 GV/c for the MDR assumes a field of 4.0 kgauss and a measurement precision of the position of $4\mu\text{m}$.]

♣ Imaging Calorimeter: the measured contamination of electrons on pions and of pion on electrons at 40 GeV/c is less than 10^{-4} at 90% efficiency for the selected particle. The energy resolution for the electrons is better than 5% up to 120 GeV.

♣ The TRD gives a separation of electrons from pions better than 10% at 90% efficiency from 2 up to 40 GeV/c.

With this performance the instrument assures that the objectives of the experiments can be reached for all the above quoted items. The expected measurement ranges for antiprotons and positrons are those reported in the fig. 14 of the lesson of A. Morselli, and the sensitivity in the antihelium to helium ratio is shown in figure 10 of the *Balloon experiments* lesson).

The final PAMELA instrument will be calibrated at CERN/SPS in summer of next year, and ready to be integrated on the RESURS-DK1 satellite in September 2002. The launch is scheduled for December 2002 from the Baikonur cosmodrome.

THE ALPHA MAGNETIC SPECTROMETER, A PARTICLE PHYSICS EXPERIMENT IN SPACE

Roberto Battiston
*Dipartimento di Fisica dell' Università and Sezione INFN
Via Pascoli, 06100 Perugia, Italy*

ABSTRACT

The Alpha Magnetic Spectrometer (AMS) is a state of the art detector for the extraterrestrial study of matter, antimatter and missing matter. During the STS-91 precursor flight in may 1998 AMS collected nearly 100 millions of Cosmic Rays on Low Earth Orbit, measuring with high accuracy their composition. We present results on the flux of proton, electron, positron and helium. Analysis of the under cutoff spectra indicates the existence of a new type of belts of energetic trapped particles characterized by a dominance of positrons versus electrons.

1 Introduction

The disappearance of the antimatter ^{1, 2, 3)} and the presence at all scales in our universe of a non luminous components of matter (dark matter) ^{4, 5)}

are two of the most intriguing mysteries in our current understanding of the structure of the Universe.

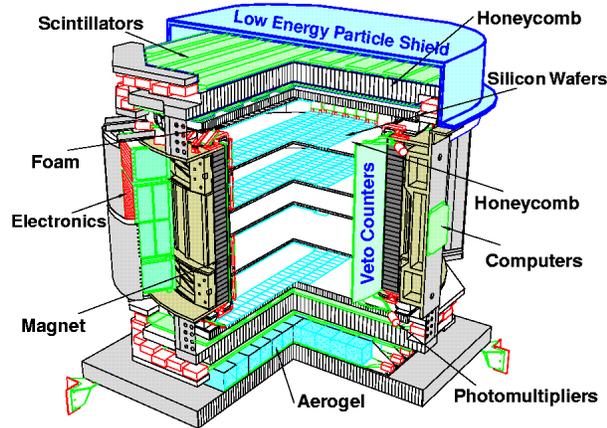


Figure 1: *AMS on the Discovery STS 91 precursor flight, June 1998.*

To study these problems, a high energy physics experiment, the Alpha Magnetic Spectrometer (AMS) ⁶⁾, is scheduled for installation on the International Space Station in 2004. Goal of AMS is to perform a three year long measurement, with the highest accuracy, of the composition of Cosmic Rays in the rigidity range 0,1 GV to several TV. In preparation for this long duration mission AMS flew a ten days precursor mission on board of the space shuttle Discovery mission STS-91 in June 1998. This high statistics measurement of CR in space, enabled, for the first time, the systematic study the behaviour of primary CR near Earth in the rigidity interval from 0,1 GV to 200 GV, at all longitudes and latitudes up to $\pm 51.7^\circ$. In this paper we present some relevant results obtained by AMS during the precursor mission. We also report the observation of high energy radiation belts in the near Earth region and on their composition, which shows remarkable differences with previously observed belts of trapped particles around our planet.

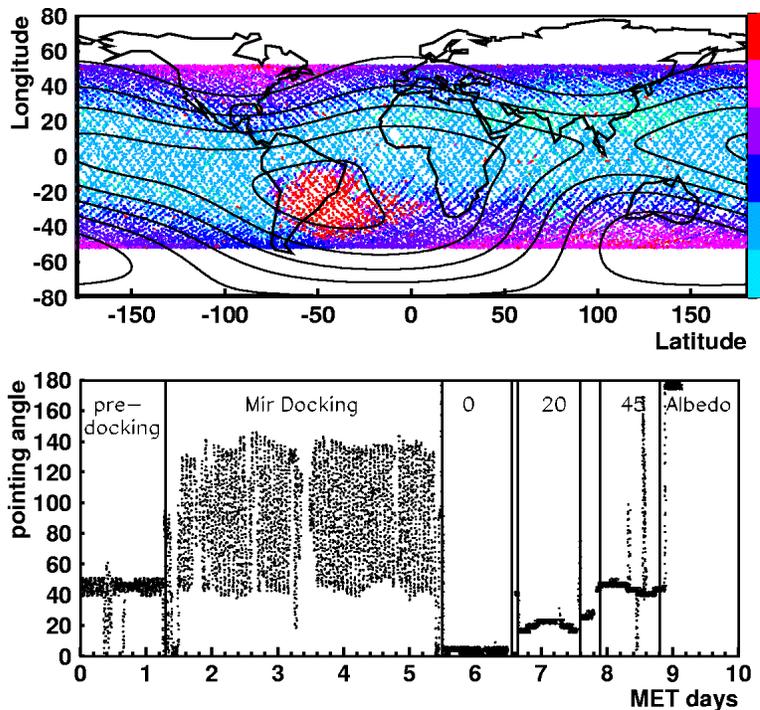


Figure 2: (a) C.R. rates versus shuttle orbits (b) and shuttle attitudes during the STS91 mission as a function of the Mission Elapsed Time (MET).

2 The AMS experiment on the STS-91 mission

Search of antimatter requires the capability to identify with the highest degree of confidence, the mass of particle traversing the experiment together with the absolute value and the sign of its electric charge.

The AMS configuration flown in 1998 on the Shuttle Discovery (Fig.1) includes a permanent Magnet, Anticounter (ACC) and Time of Flight (ToF) scintillator systems, a large area, high accuracy Silicon Tracker and an Aerogel Threshold Cherenkov counter. The magnet is based on recent advancements in permanent magnetic material and technology which make it possible to use very high grade Nd-Fe-B to build a permanent magnet with $BL^2 = 0.15 \text{ Tm}^2$ weighting ≤ 2 tons. A charged particle traversing the spectrometer triggers the experiment through the ToF system, which measures the particle velocity with a resolution of $\sim 120 \text{ ps}$ over a distance of $\sim 1.4 \text{ m}$ [11]).

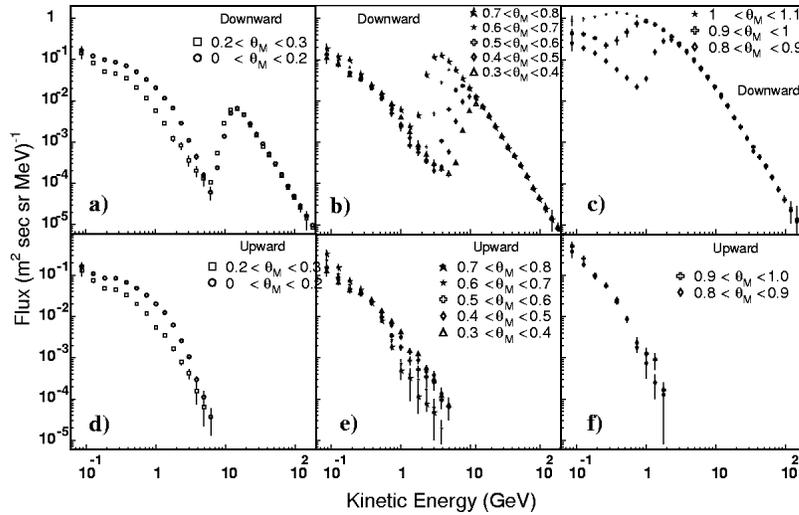


Figure 3: *Proton spectra measured by AMS for different geomagnetic latitude intervals.*

The pattern recognition and tracking is performed using the large area ($\sim 7 \text{ m}^2$), high accuracy Silicon Tracker ^{7, 9)}, which, for the Space Station mission, will be covered with 2300, high purity, double sided, $300 \mu\text{m}$ thick silicon wafers ¹²⁾, following the technology developed in Italy by INFN for the Aleph ¹⁰⁾ and L3 ⁸⁾ vertex detectors at LEP. The active area of the AMS Silicon Tracker is about an order of magnitude larger than in the case of the microstrip silicon detectors presently installed at high energy Colliders. AMS is the first high energy spectrometer based only a high precision multilayer Silicon Tracker.

The momentum resolution for AMS on the precursor mission was about ($\frac{\Delta p}{p} \sim 7\%$) at 10 GV , reaching ($\frac{\Delta p}{p} \sim 100\%$) at about 500 GV .

Four ToF scintillators layers and up to eight Silicon Tracker layers measure $\frac{dE}{dx}$, allowing a multiple determination of the absolute value of the particle charge.

By combining the various measurement it is then possible to determine the type of particle traversing the magnet and identify interesting particles with a background rejection which for anti-matter searches is expected to reach one part in 10 billions.

During the period June 2nd to June 12th, 1998 the Shuttle Discovery has performed 154 orbits at an inclination 51.7° and at an altitude varying between 390 to 350 km. During the mission AMS collected a total of about 100 Million triggers, at various Shuttle attitudes (Fig.2). In the Figure one notices the period of Shuttle to Mir docking when the Shuttle attitudes are rapidly changing with time.

Almost all results published so far (13, 14, 15) were obtained with data collected during well defined attitude periods with AMS pointing at 0°, 20° and 45° with respect to zenith (deep space).

These data are the first high quality CR data collected with a magnetic spectrometer located outside the atmosphere. The measurements cover all geomagnetic longitudes and most latitudes. These data allow a direct and accurate measurement of the CR composition and spectra, as well as a systematic study of the effects of the geomagnetic field.

The measurement of the proton flux as a function of the geomagnetic latitude (Fig. 3a-3c), shows that, in addition to the primary CR spectrum visible above the geomagnetic cutoff, there is a substantial second spectrum, extending to much lower energy and exhibiting some significant latitude dependence close to the equator. These particles cannot come from the deep space, they are on forbidden orbits, but are produced in the interaction of the primary CR with the top layers of the atmosphere. A characteristic of the second spectrum is that it is up-down symmetric (Fig. 3d-3f).

Second spectra with similar geomagnetic latitude dependence have been detected by AMS in the low energy region of the spectra of e^- , e^+ (16), D (17) and, although with weaker intensity, 3He . Fig. 4 and 5 show the results for electrons, positrons and Helium. These results on the second spectra are discussed later in the paper.

Adding all data collected above the geomagnetic cutoff it is possible to obtain a precise estimate of the primary CR differential flux. Parametrizing the omnidirectional CR flux as $\Phi(R) = \Phi_0 R^\gamma$ (R in GV) we obtain the results reported in Table 1.

It is interesting to compare the AMS measurement of the primary fluxes with previous results obtained with stratospheric balloons (18, 19, 20, 21, 22). Fig.7 shows the comparison for the proton spectrum, multiplied by $E^{2.5}$. The improved statistical significance and the wider energy interval covered by AMS

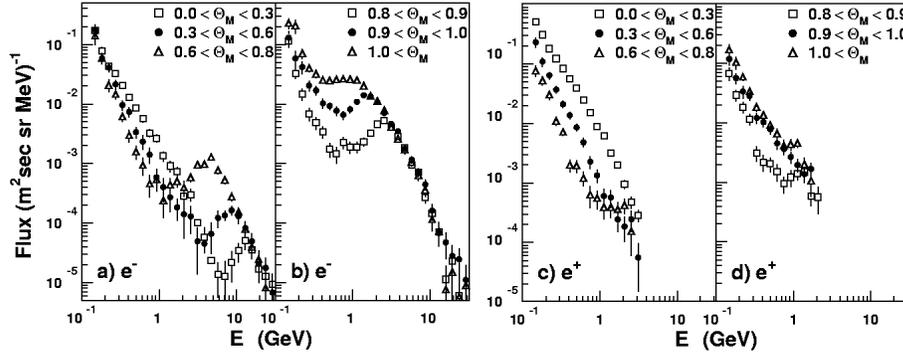


Figure 4: *Electrons and positron spectra measured by AMS during the STS91 flight.*

Table 1: *AMS results on the parametrization of proton and helium primary flux.*

Proton flux	
γ	2.78 ± 0.009 (fit) ± 0.019 (syst)
Φ_o	17.1 ± 0.15 (fit) ± 1.3 (syst) $\pm 1.5(\gamma)GV^{2.78}(m^2 s sr MeV)^{-1}$
Helium flux	
γ	2.740 ± 0.010 (fit) ± 0.016 (syst)
Φ_o	2.52 ± 0.09 (fit) ± 0.13 (syst) $\pm 0.14(\gamma)GV^{2.78}(m^2 s sr MeV)^{-1}$

data is evident: thanks to the improved accuracy obtained with only few days in space, it is possible to clarify the situation resulting from the data published over the last 15 years by the various Collaborations using different implementations of the NASA New-Mexico spectrometer ^{19, 20, 21, 22)} and by the Bess Collaboration ¹⁸⁾.

Similar consideration apply for the comparison of the measurement of Helium primary flux (Fig.7).

Both for protons as well as for Helium, AMS show a nice agreement with the measurement of the Bess Collaboration ¹⁸⁾, although our data have a smaller statistical error and extends over a wider energy interval. Using the

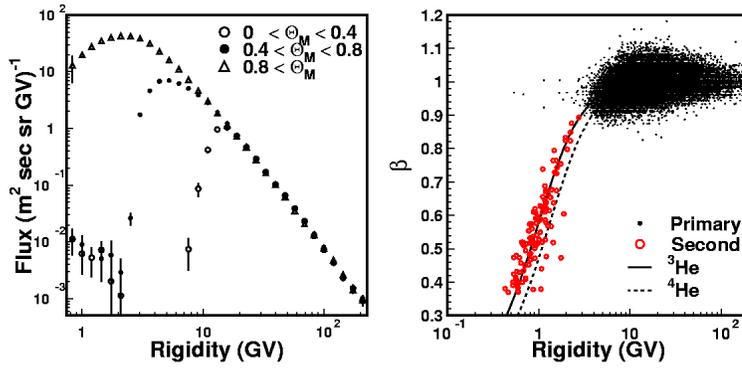


Figure 5: Helium spectra measured by AMS during the STS91 flight.

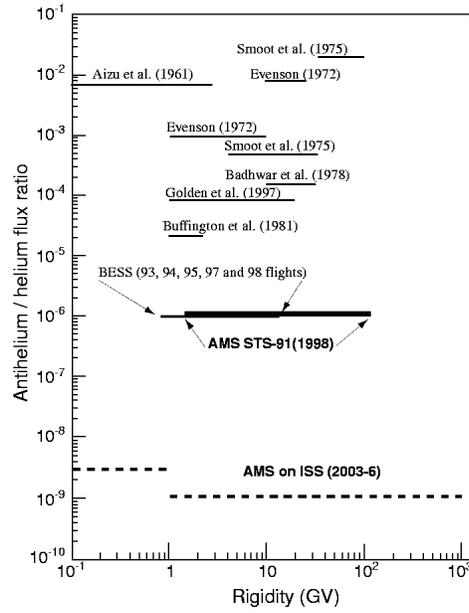


Figure 6: Antimatter limits.

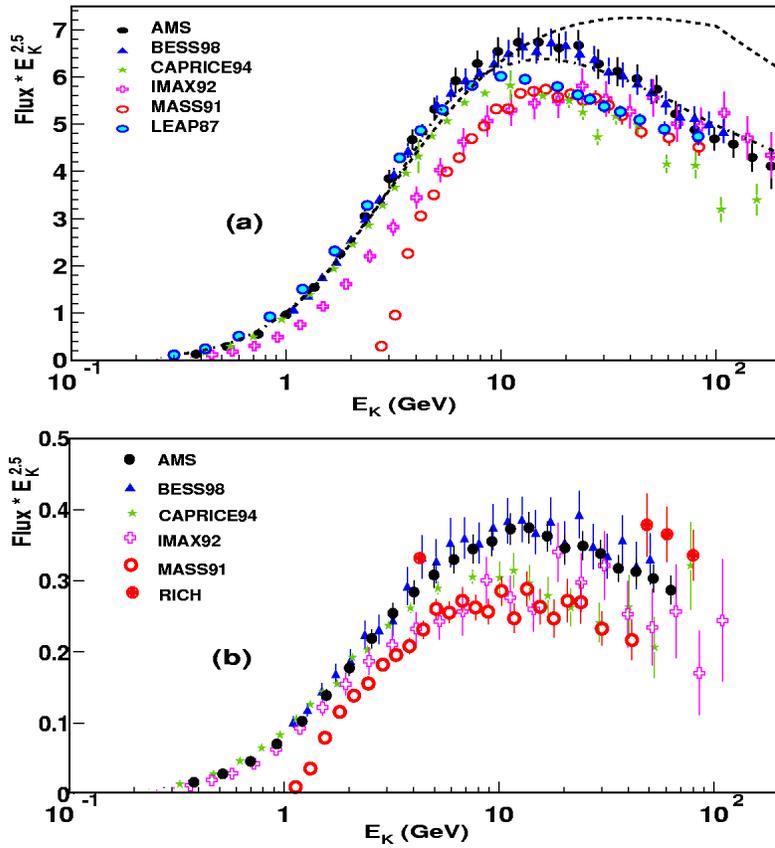


Figure 7: (a) Primary proton flux measured by AMS and compared with existing balloons measurements. The lines are parametrizations of the primary cosmic rays used in atmospheric ν flux calculation: dashed line HPPK ²³⁾, dot-dashed line Bartol group ²⁴⁾; (b) primary He flux measured by AMS and compared with existing balloons measurements.

Figure 8: *Motion of charged particles in the geomagnetic belts. A) gyration B) bouncing C) drift.*

large He sample collected by AMS a search for anti-He candidates has also been performed. Within 2.3 Millions He events no anti-He candidates have been found, up to a rigidity of 140 GV.

Assuming identical He and anti-He spectra we obtain a model independent upper limit of 1.110^{-6} over the rigidity interval 1 to 140 GV, which can be compared to previous results (Fig.7).

3 Observation of high energy particle belts

The trapping of charged particles in the quasi dipolar earth magnetic field is a classical problem, which has been studied in great detail ²⁷⁾ following Van Allen observations in 1958 ²⁹⁾. The basic physical mechanism is well understood. For sufficiently low rigidities, the trapped particles spiralize along orbits defining shells surrounding our planet.

These shells are shaped along the magnetic field lines and are roughly symmetric in latitude with respect to the geomagnetic equator (Fig.8). The motion of a trapped particle can be separated in three components, the revolution around the guiding center or gyration, the bouncing between mirror points located \approx symmetrically with respect to the geomagnetic equator (magnetic bottle), and a longitudinal drift around the earth. The geometrical locations

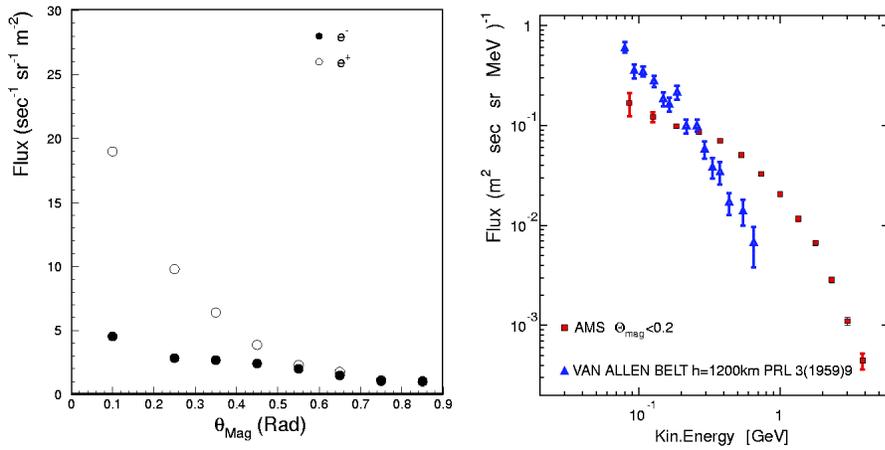


Figure 9: (a) $\frac{e^+}{e^-}$ ratio inside the belts observed by AMS, as a function of the geomagnetic latitude; (b) comparison among a typical Van Allen belt proton spectrum and equatorial AMS belts proton spectrum.

defined by the orbits of trapped particles are called shells. A shell can be univocally determined by two parameters. For example a pair of variables are L , the distance of the shell at the equator measured in unit of the Earth radius (R_{\oplus}), and B_{mir} , the value of the magnetic field at the point where the particles reverse their motion (mirror point)³⁰. Depending on the shell, B_{mir} can be locally very deep the atmosphere (it can be below the earth crust). Shells which are characterized by these value of B_{mir} cannot trap the particles, since they are lost within one or few bounces across the magnetic equator.

A particle belonging to a shell will remain on the same shell until it is disturbed by (a) interaction with the top layers of the atmosphere or other particles or (b) interaction with electrical or magnetic variable field.

Conversely, primary cosmic rays coming from deep space cannot enter a shell unless their trajectories are disturbed by some interaction with matter or fields. The existence of the shells is the result of the equilibrium between two mechanisms: some contributing to fill the shells with new particles and others removing some of the trapped particles.

Table 2: *Different types of particle belts around the Earth.*

Belt type	Particle type	Rigidity [MeV/n]	Filling mechanisms	L	Residence time [d]
Van Allen (inner)	p e^-	0.1 – 100 0.01 – 1	$n \rightarrow pe^- \bar{\nu}_e$, external belts	< 2.5	10 – 1000
Van Allen (outer)	e^- p	1 – 10 0.1 – 1	solar wind	> 2.5	1 – 10
SAMPEX	N^{+x}, O^{+x} , Ne^{+x}	10 10 – 100	Anomalous CR	2	10 – 100
AMS	p e^- e^+ 3He	100 – 1000 100 – 1000 100 – 1000 100 – 1000	primary CR interacting with the <i>atmosphere</i>	≤ 1.15	$10^{-6} - 10^{-4}$

If the dynamics of the particles trapped is well understood, the mechanisms contributing to shell stability are much less understood. They involve: interaction of high energy CR with the atmosphere creating neutrons which decays in flight, $n \rightarrow p + e^- + \bar{\nu}_e + 782KeV$, filling the belts (CRAND mechanism ²⁸), instabilities due to solar storms, as well as other types of magnetic and electric instabilities. It should be pointed out, however, that the mechanisms proposed are compatible with the observed dominance of protons and electrons in the Van Allen belts.

The shell can be classified by their composition and location. The original Van Allen belts contain only proton and electrons and extend to very large distance from the earth, up to $L \approx 6$. Van Allen belts are divided into inner and outer belts, since there is a dip in the particle flux intensity at about $2.5 L$. During the last 20 years, there have been reports of the observation of a low flux of trapped ions, mainly *He* and *O*, with traces of *C* e *N*, and having energies of a few MeV/n and $L = 3 - 4$. These particles are extracted from the upper layers of the atmosphere during solar storms. More recently, nearly 40 years after Van Allen discovery, the analysis of SAMPEX data ³²) has shown the existence of belts included in the inner Van Allen belts, containing heavier nuclei like *N*, *O*, *Ne* with rigidities of the order of $10/MeV$.

The SAMPEX belts are different from the Van Allen belts mainly because of their composition due to a different the filling mechanism, which is likely

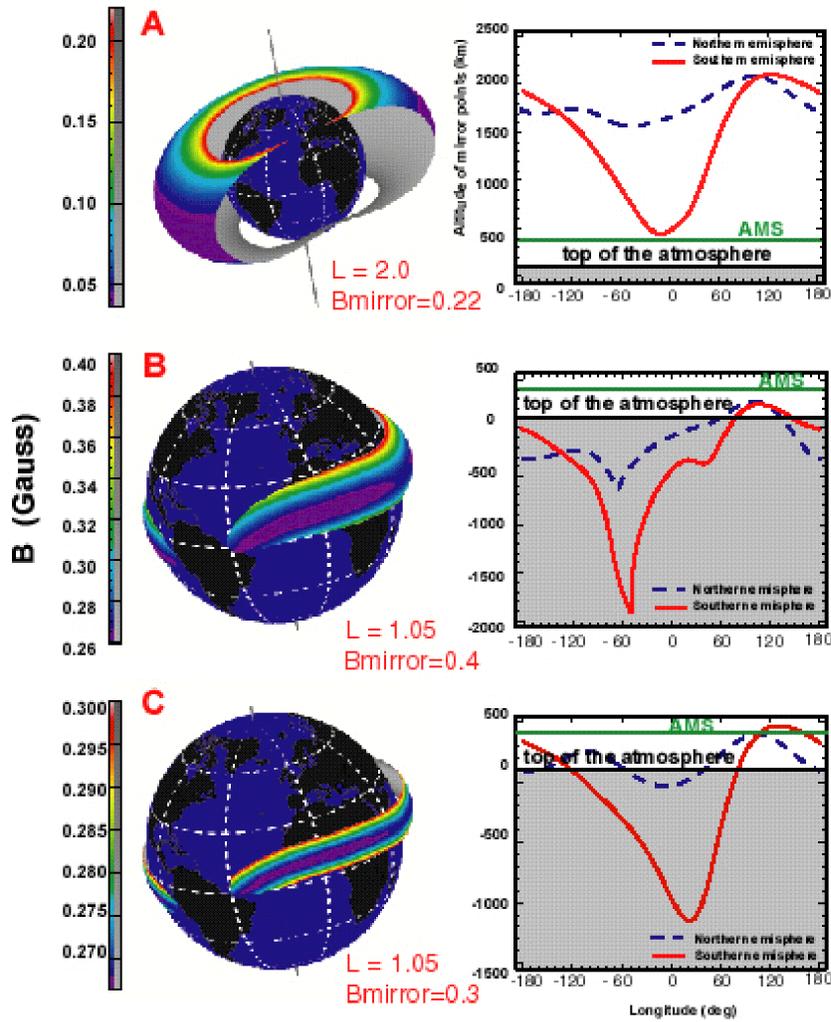


Figure 10: Van Allen versus AMS belts. (A) Van Allen belts have high L values, B_{mirror} is located mainly above AMS orbits, particles weakly interact with the atmosphere and have lifetimes ranging from days to months. AMS belts have $L \leq 1.5$, and depending on the value of the B_{mirror} their lifetime ranges from fractions of a second (B) to several seconds (C).

due to the interaction of the so called Anomalous Cosmic Rays with the Earth atmosphere 33, 31).

The belts observed by AMS are different in composition since they also contain a large fraction of positrons, but also deuterium and ${}^3\text{He}$. These particles have not been observed in the Van Allen or SAMPEX belts. Particularly striking is the abundance of positrons versus electrons (Fig.4), with a ratio exceeding a factor of four in the equatorial region (Fig. 9a).

AMS observed shells with $L \leq 1.15$, well below the inner Van Allen belts. In the belts studied by AMS the observed proton spectrum is harder (Fig. 9b) than in the case of Van Allen belts. This can be understood since their location is closer to the earth and the particles do experience a stronger trapping field. Another difference with the Van Allen belts is the residence time of the trapped particles, computed using computer based tracing techniques, which is in the region of seconds and not days or weeks. These shells cannot be observed by stratospheric balloons, since their mirror fields are above the atmosphere except in correspondence of the South Atlantic Anomaly. It follows that the observed particles do not belong to the various types of albedo particles reported in the past by experiments on balloons.

In Table 2 we summarize the main features of the different type of belts identified during the last 40 years. As we can see the situation is very varied, corresponding to different filling mechanisms. Since we are dealing with continuous distributions, the reported intervals (rigidity, L, residence time) should be taken as typical order of magnitudes. In Fig.10, we compare the structure of the AMS belts to the Van Allen belts as well as the dependence of the mirror field altitude with the longitude.

4 Conclusions

The first mission of the Alpha Magnetic Spectrometer, although lasting only ten days, has been scientifically very rewarding, allowing for the first time a very detailed measurement of high energy cosmic rays outside the atmosphere. In addition to the most accurate measurements obtained so far for the primary flux of $p, e^+, e^-, D, {}^3\text{He}$ and ${}^4\text{He}$ spectra over most of the earth surface, these results have shown the existence of a substantial second spectrum of high energy particles trapped within low altitude belts. These new belts have a very characteristic composition, dominated by positively charged particles, mainly

Table 3: *Physics capabilities of AMS after three years on the ISS*

Elements	Sensitivity	(Now)	Energy Range(GV)	Physics
e^+	10^8	($\sim 10^3$)	0.1 – 100	↑
\bar{p}	500000	(~ 30)	0.5 – 100	Dark Matter
γ			0.1 – 300	↓
He/He	$\frac{1}{10^9}$	($\frac{1}{10^5}$)	0.5 – 20	Antimatter
C/C	$\frac{1}{10^8}$	($\frac{1}{10^4}$)	0.5 – 20	CP, GUT, EW
D, H_2	10^9		1.0 – 3.0	↑
${}^3He/{}^4He$	10^9		1.0 – 3.0	Astrophysics
${}^{10}Be/{}^9Be$	2%		1.0 – 3.0	↓

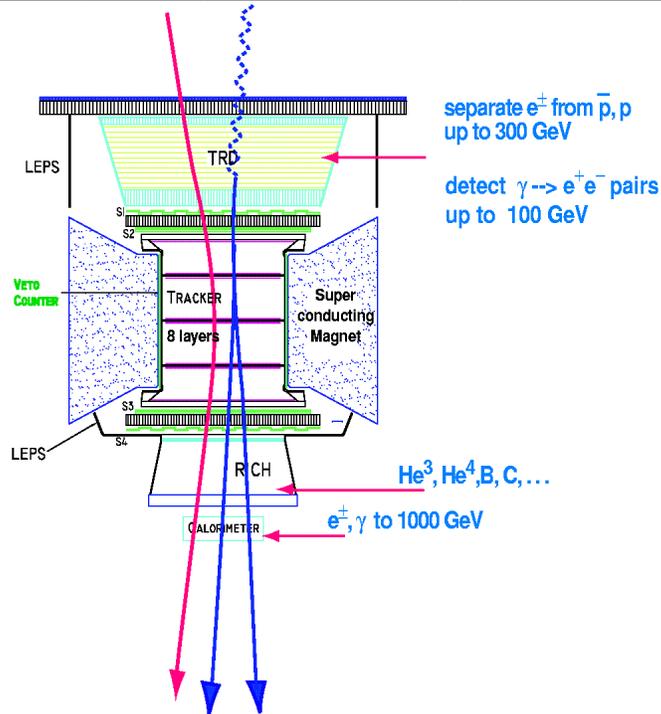


Figure 11: *Configuration of AMS on the ISS for the three years mission scheduled on UF4 in 2004.*

p, e^+ and D . Their existence should be taken into account when calculating radiation doses for astronauts on the ISS or background rates for low orbit satellites.

AMS is currently being refurbished to be ready for a three years mission with UF4 in 2004. A stronger magnetic field from a superconducting magnet, $B = 0,7 T$, a fully equipped Silicon Tracker, together with three powerful particle identification detectors, a Transition Radiation Detector, a Ring Imaging Cherenkov (RICH) detector and an Electromagnetic Calorimeter, will allow precise particle identification up to $O(TeV)$ of energy (Fig.11). The physics capabilities of AMS after three years of exposure on the ISS are summarized in Table 3. AMS will be the only large acceptance magnetic facility which will be exposed for long time in space. It will allow a measurements of the flux and composition of Cosmic Rays with an accuracy orders of magnitude better than before. The large improvement in sensitivity given by this new instrument, will allow us to enter into a totally new domain to explore the unknown.

5 Acknowledgment

This work has been partially supported by the Italian Space Agency (ASI) under contract ARS-98/47.

References

1. Steigmann, G., Ann. Rev. Astron. Astroph., **14** p. 339, 1976.
2. Kolb, E.W., Turner, M.S., Ann. Rev. Nucl. Part. Sci. **33** p. 645, 1983.
3. Peebles, P.J.E., Principles of Physical Cosmology, Princeton University Press, Princeton N.J., 1993.
4. Ellis, J. et al., Phys. Lett. **B214**, p. 403, 1988.
5. Turner, M.S., Wilzek, F., Phys. Rev. **D42**, p. 1001, 1990.
6. Ahlen, S. et al., Nucl. Inst. Meth. **A350**, p. 351, 1994.
7. Battiston, R., Nucl. Instr. Meth. (Proc. Suppl.) **B44**, p. 274, 1995.
8. Acciarri, M. et al., Nucl. Inst. Meth. **A289** p. 351-355, 1990.

9. Alcaraz, J. et al., *Il Nuovo Cimento* **112A**, p. 1325-1344, 1999.
10. Batignani, G. et al., *Nucl. Inst. Meth.* **A277** p. 147, 1989.
11. Alvisi, D. et al., *Nucl. Inst. Meth.* **A437** p. 212, 1999.
12. Produced at CSEM, SA Rue J. Duot 1, P.O. Box, CH-2007 Neuchatel, <http://www.csem.ch>.
13. AMS Collaboration, Alcaraz, J. et. al., Search for Antihelium in Cosmic Rays, *Phys. Lett.* **B461**, p. 287, 2000.
14. AMS Collaboration, Alcaraz, J. et. al., Protons in Near Earth Orbit, *Phys. Lett.* **B472**, p. 215, 2000.
15. AMS Collaboration, Alcaraz, J. et. al., Leptons in Near Earth Orbit, *Phys. Lett.* **B484**, p. 10, 2000.
16. AMS Collaboration, Alcaraz, J. et. al., Cosmic Protons, *Phys. Lett.* **B490**, p. 27, 2000.
17. Lamanna, G. PhD Thesis, University of Perugia, October 2000, unpublished.
18. BESS98, Sanuki, T. et al., *astro-ph/0002481*, 2000.
19. CAPRICE94, Boezio, M. et al., *ApJ* **518**, p. 457, 1999.
20. IMAX92, Menn, W. et al., *The Astroph. J.* **533**, p. 281, 2000.
21. MASS91, Bellotti, R. et al., *Phys. Rev.* **D60**, p. 052002, 1999.
22. LEAP87, Seo, E.S. et al., *ApJ* **378**, p. 763, 1991.
23. HPPK, Honda, M. et al., *Phys. Rev.* **52**, p. 4985, 1995.
24. BARTOL, Lipari, P. and Stanev, T., Talk given at Now 2000 Conference, 2000.
25. Smoot, G.F. et al., *Phys. Rev. Lett.* **35** p. 258-261, 1975; Steigman, G. et al., *Ann. Rev. Astr. Ap.* **14** p.399, 1976; Badhwar, G. et al., *Nature* **274** p. 137, 1978; Buffington, A. et al., *ApJ* **248** p. 1179-1193, 1981; Golden, R.L. et al., *ApJ* **479** p. 992, 1997; Ormes, J.F. et al., *ApJL* **482** p. 187, 1997; Saeki, T. et al., *Phys. Lett.* **B422** p. 319, 1998.

26. Nozaki, M., OG.1.1.23, 26th ICRC, Salt Lake City, Utah, 1999.
27. For a recent review see Walt, M., Radiation Belts Models and Standards, AGU Geophysical Monograph 97, p.1, 1997.
28. Singer, S.F., Phys. Rev. Lett. **1** p. 181, 1958; Hess, W.N., Phys. Rev. Lett. **3** p.11, 1959; Kellogg, P., J. Geophys. Res. **65** p. 2,705, 1960; Vernov, S.N. et al., Soviet Physics, Doklady **4** p.154, 1959.
29. Van Allen, Ludwig, Ray, and McIlwain, IGY Satellite Series Number **3**, **73** Natl. Acad. Sci., Washington D.C., 1958; Van Allen, McIlwain, and Ludwig, J. Geophys. Research **64**, p. 271, 1959; Van Allen, J.A. and Frank. L.A., Nature **183** p. 430, 1959.
30. McIlwain, C.E., J. Geophys. Res. **66**p. 3681-3691, 1961.
31. Mewaldt, R.A., Radiation Belts Models and Standards, AGU Geophysical Monograph 97, p.35, 1997.
32. Cook, W.R., IEEE Trans. Geosci. Remote Sensing **31** p. 557-564, 1993.
33. Cummings, J.R. et al., Geophys. Res. Lett. **20** p. 2003-2006, 1993; Cummings, J.R. et al., IEEE Trans. Nucl. Sci. **40** p. 1459-1462, 1993.

EUSO: USING HIGH ENERGY COSMIC RAYS AND NEUTRINOS AS MESSENGERS FROM THE UNKNOWN UNIVERSE ^(^o)

L. Scarsi

Istituto di Fisica Cosmica e Informatica / CNR and University of Palermo,

ABSTRACT

Exploiting the Earth Atmosphere as a giant detector for the incoming extraterrestrial flux of High Energy Cosmic Rays and Cosmic Neutrinos, the mission " EUSO - Extreme Universe Space Observatory " is devoted to the exploration of the domain of the highest energy processes occurring in the Universe up to its accessible boundaries. The observable is provided by the Air Nitrogen fluorescence light emitted in the UV band 300 – 400 nm by the Extensive Air Showers produced by the cascading processes of the Primary C.R. Particles interacting with the Atmosphere. The EUSO telescope is based on a double Fresnel lens optics (diameter 2.5 m) coupled to an highly pixelized focal surface composed by multianode PMTs ; the image at the Earth surface is detailed at 1 Km² over a total of several hundred thousand of Km². EUSO will fly on the International Space Station accommodated as External Payload of the European Space Agency Columbus module . The mission is scheduled to last 3 years, with the start of operations foreseen for 2007/8 . The expectations are of a collection rate of a thousand events / year for Cosmic Rays at $E > 10^{20}$ eV together with tens / hundreds Cosmic Neutrinos at energy above about 4×10^{19} eV. EUSO is the result of the collaborative effort of several Institutions in Europe, Japan and USA and it is conceived within the science program sponsored by various Space Agencies coordinated by ESA

^(^o)This text is largely derived from contributions made by the Author in 2000/2001 to other Workshops and Conferences and from documentation submitted by the EUSO Consortium to the European Space Agency as a part of the proposal EUSO

1 Introduction

The Cosmic Radiation can be considered the "Particle channel" complementing the "Electromagnetic Channel" proper of the conventional Astronomy.

A classic presentation of the Cosmic Ray Energy Spectrum is shown in Fig.1; an unconventional view (which I borrowed from a colleague of Karlsruhe, where it was first shown at the Cerimonial organized in honor of Dr. Shatz) is given in Fig.2 to illustrate in

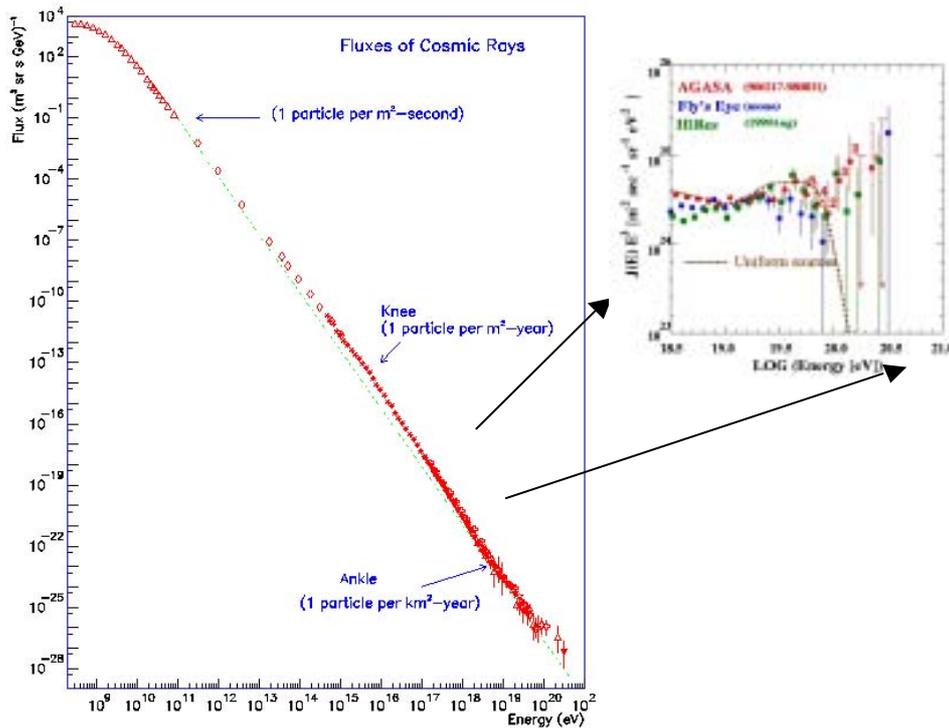


Figure 1. The observed cosmic ray spectrum for $E > 10^8$ eV showing the principle features. The inset shows the high-energy part with the overall E^{-3} dependence removed as observed by AGASA (Takeda et al. 1998), Fly's Eye and Hires (Teshima 2000). The dashed line shows the effect of the GZK cutoff assuming a homogenous source population filling the Universe. The numbers are the actual number of events in each bin, made in the knowledge of the nature of Cosmic Rays of the relatively modest energies (up to the "knee" at 10^{14} - 10^{15} eV); the Cosmic Radiation on the higher energy side on the other hand presents us with the challenge of understanding its origin and its connection with fundamental problems in Cosmology and Astroparticle Physics.

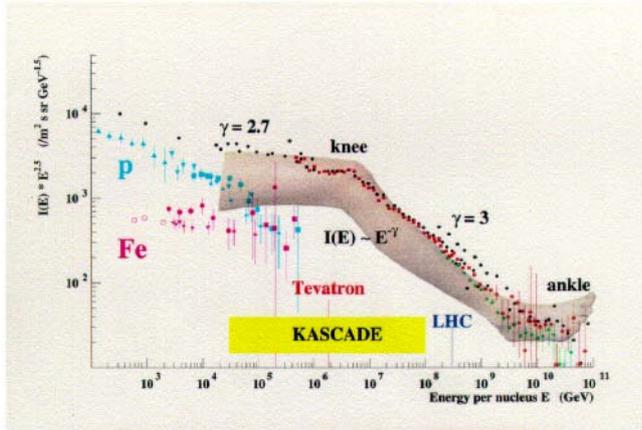


Figure 2. The Cosmic Ray energy spectrum: anthropomorphic presentation.

an anthropomorphic perspective the features conventionally nominated “knee“ (around 10^{15} eV) and “ankle“ (above 5×10^{18} eV). The remarkable “feminine leg“ in the figure is that of the famous German movie Star Marlene Dietrich. Today substantial progresses have been Focal points are represented by:

- i) The change in the spectral index at $\sim 5 \cdot 10^{18}$ eV (“Ankle “)
- ii) Existence of "Cosmic Rays" with energy $E > 10^{20}$ eV: (EECR) (Fig.1). A direct question arising is: what is the maximum Cosmic Ray energy, if there is any limit? Addressing the theoretical issue concerning the production and propagation of 10^{20} eV Primary quanta is problematic and it involves processes still little known

2 The Universe and the probing depth of the Extreme Energy Cosmic Radiation.

(From the document “Report on the Accommodation of EUSO on the Columbus Exposed Payload Facility: ESA/MSM-GU/2000.462/AP/RDA. December 2000).

Soon after the discovery of the cosmic microwave background radiation (CMB) by Penzias and Wilson in 1965, Greisen, and, independently, Zatsepin and Kuzmin pointed out that this radiation would make the universe opaque to cosmic rays of sufficiently high energy. For protons, e.g., this occurs when the pion production threshold is reached (about 5×10^{19} eV, if the Lorentz transformations of Relativity still hold at $\gamma \geq 10^{11}$). The reaction $p + \gamma \rightarrow \Delta^+ \rightarrow p + \pi^0$ or $n + \pi^+$ will lead to an effective attenuation length of 50 Mpc for a proton of 10^{20} eV. This is about the size of the Virgo cluster to which our galaxy belongs, and is just a small fraction of the size of the Universe. Table 1 summarizes the effects introduced on the primary EECR components, including gamma rays, by the interaction with the CMB. To a much lesser extent, neutrinos decay above 10^{21} - 10^{22} eV by Z_0 -resonance with cosmic neutrino background (CNB).

Table 1. Extreme energy processes that cutoff the energy spectrum of particles in Universe.

	Process	Cutoff Energy	Mean free path
Protons	$p + \gamma_{2.7K} \rightarrow \pi^0 + X$	eV	50 Mpc
Nuclei	$A + \gamma_{2.7K} \rightarrow \Delta^{++} + X$	$\geq 5 \times 10^{18}$ eV/n	100 Mpc
Gamma-rays	$\gamma + \gamma_{2.7K}$	$\geq 10^{14}$ eV (at 10^{20} eV)	10 Mpc (at 10^{20} eV)
Neutrinos	$\nu + \nu_{1.95K} \rightarrow (W/Z_0) + X$	$\geq 4 \times 10^{22}$ eV	40 Gpc

The Greisen-Zatsepin-Kuzmin effect shapes in a complicate way the form of the observed energy spectrum of the EECR as a function of the distribution of the extragalactic sources in the Universe (see Figure 3).

2.1 Sources of Extreme Energy Cosmic Rays

Many sources of high-energy particles could exist in the Universe at distances beyond the GZK length of 50 Mpc. The predicted sharp cutoff of energy spectrum above 5×10^{19} eV may not be present if a possible high abundance of cosmological neutrino events, or others from nearby sources, are dominant. A significant bump should exist at 10^{19} - 10^{20} eV due to the GZK effect for protons, because cosmological protons that were accelerated to the super-

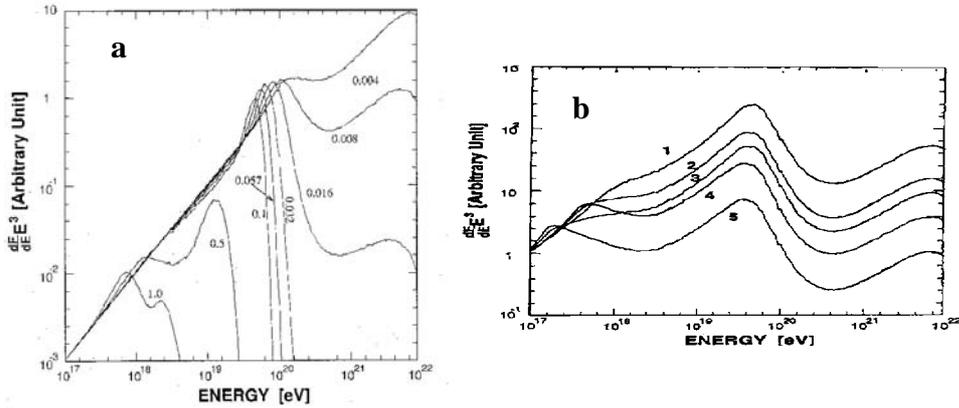


Figure 3. Left: Energy spectra from a single source of protons with an E^{-2} spectrum, for various source distances between $z = 0:004$ and 1 (i.e. between 2 and 5000 Mpc). **Right:** The nucleon spectrum at the trans-GZK and super-GZK energies predicted from different models of the cosmic evolution, including the hypothetical component at EE up to 10^{23} eV. The curve 1 is for the case of homogeneous evolution for the redshifts $0 \leq z \leq 2$. The curve 5 shows non-homogeneous distribution up to $z = 4$ in which early cosmological era has more active EE sources. Curves 2-4 are for intermediate models. Cosmological neutrinos and their secondaries are not included in this figure.

GZK energies decay down to the trans-GZK energies, and pile up at sub-GZK energies. The details of the energy spectrum in the trans-GZK energy regime ($10^{19} - 2 \times 10^{20}$ eV) depend on the model of the evolution of Universe from about 10 billion years ago (or $z \sim 5$). High event

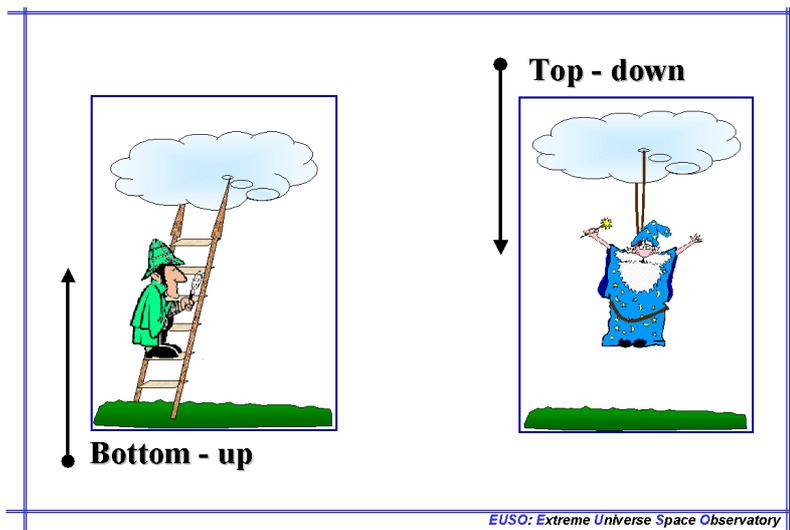
statistics can provide information on the evolution of the highest energy Universe. This information on the evolution of Universe may be distorted by the probable existence of super-GZK neutrinos. However, using EUSO the neutrino energy spectrum can be identified and directly measured, and the non-neutrino spectrum can be statistically corrected by subtracting relevant neutrinos and their secondaries. The evolution of energetic sources in universe will be examined by such statistical means.

The current summary of the data from AGASA, Fly's Eye and Hi-Res is shown in the insert of Fig. 1 and suggests that the observed cosmic ray spectrum is not necessarily cut off following the predicted GZK process for protons and nuclei. It suggests that either the majority of cosmic rays originated within the GZK length (< 50 Mpc), or the relativity principle fails at extreme energies, or some other unknown sources exists.

Concerning the distribution of arrival directions, AGASA and world wide data summary of cosmic rays above about the GZK cutoff energy shows a quasi-isotropic distribution in the sky clearly suggesting an extragalactic origin. Among them, 6 pairs and 1 triple set of spatially correlated events within 2 years from 58 events were recognized by AGASA only, while the world-wide data show 9 pairs and two triples. The chance coincidence probability for these "clusters" of events is less than 0.07%, and therefore, the particles of a "cluster" possibly had the same sources

Although the existence of the highest energy cosmic rays is proven, their origin is still an enigma despite the efforts of many theorists and experimentalists. One is led to the conclusion that they have an entirely different origin than the lower energy cosmic rays. The present data raise questions of great importance for astrophysics, cosmology, and fundamental physics.

Focusing the attention on the primary sources, the general production mechanisms proposed for the EECRs can be classified as:



BOTTOM-UP, with acceleration in rapidly evolving processes occurring in Astrophysical Objects. The scenario involves astrophysical objects such as, e.g. AGNs and AGN radio lobes. The study of these objects is, besides radio observations, a main goal of

X-ray and Gamma-ray astrophysics of the late 90's. An extreme case in this class is represented by the Gamma Ray Bursts, found to be located at cosmological distances. The observation of "direction of arrival and time" coincidences of GRBs and Extreme Energy Neutrinos ($E \geq 10^{19}$ eV) in the EUSO mission could provide a crucial test for the identification of the observed GRBs as EECR sources in spite of their location at distances well above the GZK limit.

TOP-DOWN Processes. This scenario arises from the cascading of ultrahigh energy particles from the decay of topological defects. Cosmic Strings would play an essential role for releasing the X-bosons emitting the highest energy quarks and leptons. This process could occur in the nearby Universe. The relics of an early inflationary phase in the history of the Universe may survive to the present as a part of dark matter and account for those unidentified EECR sources active within the GZK boundary limit. Their decays can give origin to the highest energy cosmic rays, either by emission of hadrons and photons, as through production of EE neutrinos.

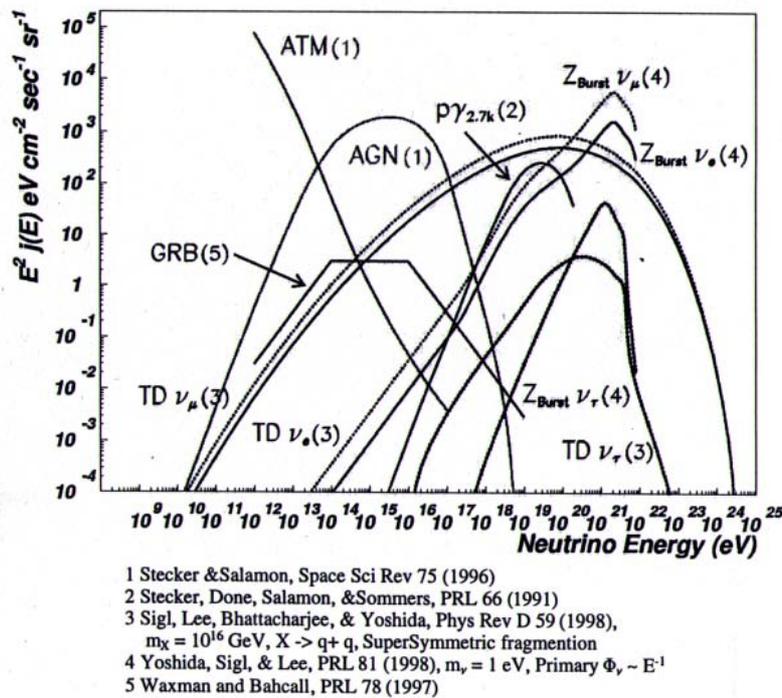


Figure 4. Predicted neutrino fluxes for various models. The large range of predicted fluxes should be noted as should the number of models which exceeds the number of events with energies $> 10^{20}$ eV !

1.2 Neutrino induced Air Showers

Neutrinos with high enough energy can produce detectable EAS observable by EUSO. This will provide precious information about their origin together with that of the EECR. Not suffering the GZK effect and being immune from magnetic field deflections, or from a delay caused by the quantum relativity effects, neutrinos are ideal for disentangling source related mechanisms from propagation related effects. The opening of the neutrino astronomy channel will allow the extreme boundaries of the Universe to be probed.

From the Astroparticle Physics point of view, the EECRs have energies only a few decades below the Grand Unification Energy (10^{24} - 10^{25} eV), although still far from the Plank Mass of 10^{28} eV.

However, neutrinos are elusive objects with a low interaction probability, to such an extent that they can be neglected as observable EAS initiators for all ground based detectors, present or planned. Even for the largest planned ground based cosmic ray detector (the Auger project), in the most optimistic case, the expected rate is only a few events per year. EUSO, with its large sensitive area and accessible mass target of the order of 10^{13} tons of atmosphere, will be sensitive to this class of events.

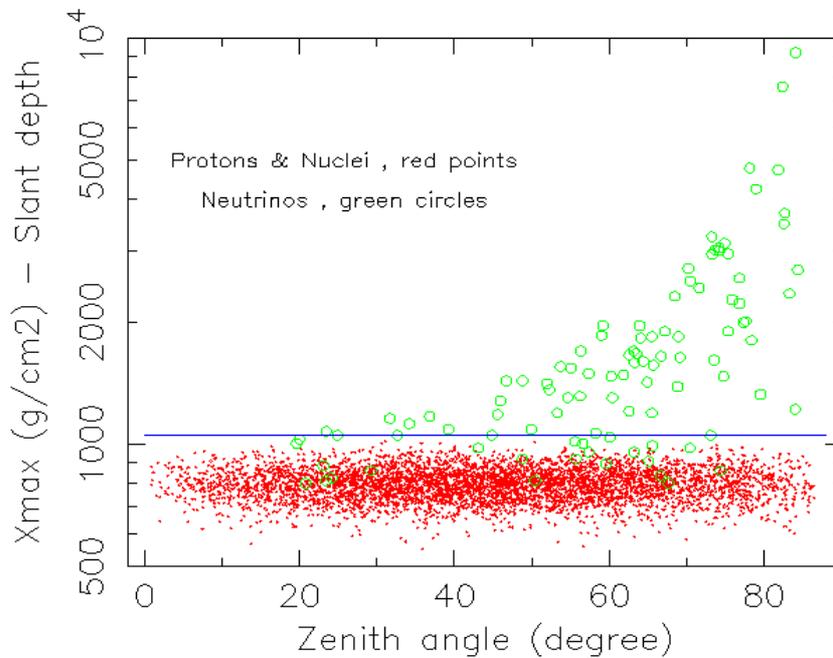


Figure 5. Shower depth distribution from Monte Carlo simulations showing how neutrino and proton and nuclei induced events can be distinguished

The expected neutrino event rate ranges from a few events per year (GZK processes, AGN, GRB sources etc) to 150 per year according to the effectiveness of the "topological

defects" hypothesis (see Fig. 4). Observationally, neutrino induced EAS can be distinguished from background events and from other EECR EAS by selecting events with large zenith angles which initiate deep in the atmosphere (Fig. 5). A nearly horizontal τ -neutrino event with an energy $>10^{19}$ eV can be identified by a "double bang" structure. Both the initial shower in the $\nu_\tau \rightarrow \tau$ interaction, and another, by the τ -decay, can be seen because of the long path length ($\sim 1000 [E/10^{20} \text{ eV}] \text{ km}$) for τ -decays.

3 Observational problems

The extremely low value for the EECR flux, corresponding to about 1 event per km^2 and century at $E > 10^{20}$ eV, and the extremely low value for the interaction cross section of neutrinos, make these components difficult to observe if not by using a detector with exceptionally high values for the effective area and target mass. The integrated exposure ($2 \cdot 10^3 \text{ km}^2 \text{ yr sr}$) available today for the ground based arrays operational over the world is sufficient only to show the "ankle" feature at $\sim 5 \cdot 10^{18}$ eV in the Cosmic Ray energy spectrum and the existence of about ten events exceeding 10^{20} eV; the limited statistics excludes the possibility of observing significant structures in the energy spectrum at higher energies. Experiments carried out by means of the new generation ground-based observatories, HiRes (fluorescence) and Auger (hybrid), will still be limited by practical difficulties connected to a relatively small collecting area ($< 10^4 \text{ km}^2 \text{ sr}$) and by a modest target mass value for neutrino detection.

To overcome these difficulties, a solution is provided by observing from space (Fig.6) the atmosphere UV fluorescence induced by the incoming extraterrestrial radiation, which allows to exploit up to millions $\text{km}^2 \text{ sr}$ for the acceptance area and up to 10^{13} tons as target for neutrino interaction. This is the philosophy of the "AirWatch Programme" and "EUSO" is a space mission developed in the AirWatch framework.

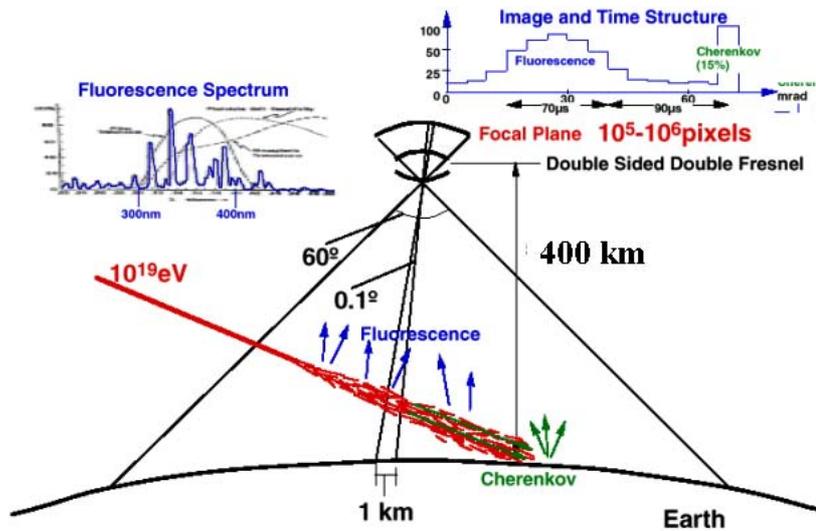


Figure 6 . Observation of EAS from Space.

The Earth atmosphere in fact constitutes the ideal detector for the Extreme Energy Cosmic Rays and the companion Cosmic Neutrinos. The EECR particles, interacting with the air nuclei, give rise to propagating Extensive Air Showers (EAS) accompanied by the isotropic emission of UltraViolet fluorescence (300-400 nm) induced in Nitrogen by the secondary charged particles in the EAS as result of a complex relativistic cascade process; an isotropically diffuse optical-UV signal is also emitted following the impact on clouds, land or sea of the Cherenkov beam accompanying the EAS. A Shower corresponding to a Primary with $E > 10^{19}$ eV forms a significant streak of fluorescence light over 10-100 km along its passage in the atmosphere, depending on the nature of the Primary, and on the pitch angle with the vertical.

Observation of this fluorescence light with a detector at distance from the shower axis is the best way to control the cascade profile of the EAS. When viewed continuously, the object moves on a straight path with the speed of light. The resulting picture of the event seen by the detector looks like a narrow track in which the recorded amount of light is proportional to the shower size at the various penetration depth in the atmosphere. From a Low Earth Orbit (LEO) space platform, the UV fluorescence induced in atmospheric Nitrogen by the incoming radiation can be monitored and studied. Other phenomena such as meteors, space debris, lightning, atmospheric flashes, can also be observed; the luminescence coming from the EAS produced by the Cosmic Ray quanta can be on the other hand disentangled from the general background exploiting its fast timing characteristic feature. EUSO observes at Nadir from an orbital height of about 400 km. It is equipped with a wide angle Fresnel optics telescope (60° full FoV) and the focal plane segmentation corresponding to about 1 km^2 pixel size on the Earth surface. The area covered on Earth is of about 160000 km^2 . Exploiting the high speed of the focal plane detector (10 ns class), EUSO is able to reconstruct the inclination of the shower track by the speed of progression of the projected image on the focal surface and to provide the tri-dimensional reconstruction of the EAS axis with a precision of a degree (or better) depending on the inclination. By measuring the EAS front luminosity with the photoelectrons (PE) detected by the MAPTs covering the focal surface, EUSO registers the longitudinal development of the EAS.

1.1 EUSO General Requirements and Main Goals

For a significant observation from a space mission the assumed values are:
a) Geometrical exposure of $(5 \cdot 10^4 - 10^5) \text{ km}^2 \text{ sr}$ considering a duty cycle of 0.1-0.15;
b) EAS energy threshold at about $5 \cdot 10^{19}$ eV.

EECR statistics. About 10^3 events/year (an order of magnitude above those expected by the presently planned ground based experiments) to allow a quantitative energy spectral definition above 10^{20} eV, together with the evidence of possible anisotropy effects and clustering (if any) for the directions of arrival.

Neutrino events. The expected event rate ranges from several events/year (AGN, GRB source) to several events/day according to the effectiveness of the "topological defects" hypothesis. From the observational point of view, the neutrino induced EAS can be distinguished from background and from other EECR EAS by triggering on horizontal showers initiating deep inside the atmosphere. Moreover neutrinos with energy of about $10^{15} - 10^{16}$ eV interacting in the solid earth and emerging upward in the atmosphere create showers which can be detected by EUSO by means of the Cherenkov beamed signal induced in the atmosphere, extending the capability of EUSO to this lower neutrino astronomy energy band. A horizontal tau-neutrino event at energies greater than 10^{19} eV can be identified by a "double bang" structure. Both the initial shower at the $\nu_\tau \rightarrow \tau$ interaction,

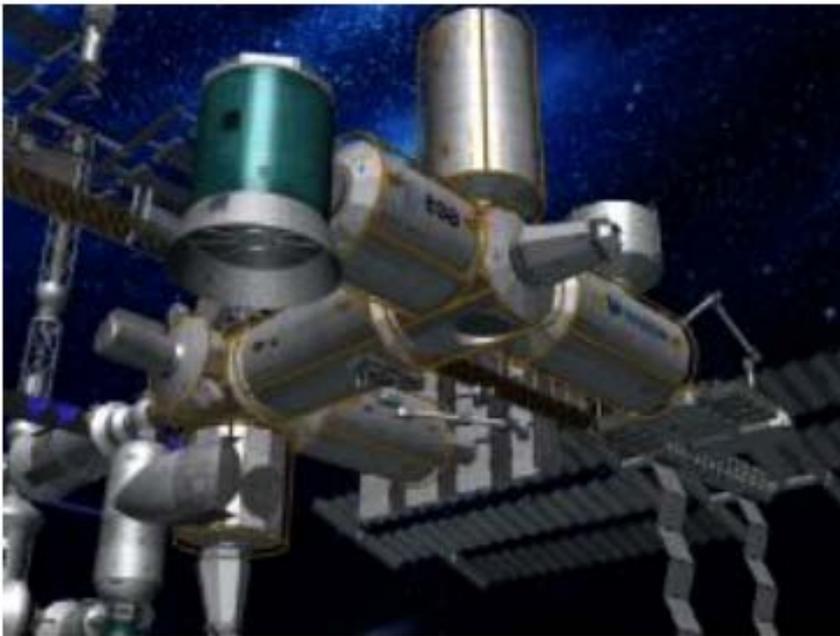
and another, by the τ -decay, can be seen because of the long enough path-length ($\sim 1000 [E/10^{20} \text{ eV}] \text{ km}$) for τ -decays observable by EUSO. Tau-neutrinos above 10^{15} eV , on the other hand, will be observed and identified as Earth-penetrating “upward” showers (by Cherenkov). High ν_τ flux by the $\nu_\mu \rightarrow \nu_\tau$ oscillation and the low detection threshold energy for them allow EUSO to make oscillation experiments in space as well as ν_τ astrophysics of AGN above 10^{15} eV .

1.2 EUSO Schematic Outline.

EUSO, originally proposed to ESA in January 2000 for a free-flyer LEO mission, has been approved in March 2000 for an “Accommodation study” on the ISS International Space Station. As a result, EUSO is now undergoing a “bridging phase” to enter “Phase A design” carried out by Industry under an ESA contract (transportation and transfer to the ISS/Columbus EPF (Fig.7)).

Under the assumption of both a LEO ($\sim 500 \text{ km}$ altitude) free-flyer mission or the ISS accommodation (400 km average altitude), the coverage of the observable atmosphere surface at the scale of thousand kilometers across and the measurement of very fast and faint phenomena like those EUSO is interested in, requires:

Figure 7. EUSO at the COF-EPF.



optical system with large collecting area (because of the faint fluorescence signal) and wide equivalent field of view covering a sizable half opening angle around the local Nadir (to reach geometrical factor of the order of $10^6 \text{ km}^2 \text{ sr}$),

focal plane detector with high segmentation (single photon counting and high pixelization), high resolving time (~10 ns), contained values for weight and power,

trigger and read-out electronics prompt, simple, efficient, modular, capable to handle hundreds of thousands of channels, and comprehensive of a sophisticated on-board image processor acting as a trigger.

1.3 EUSO Payload: The “Main Telescope”

The EUSO Main telescope is presented schematically in the artistic view of Fig. 8. The instrument consists of three main parts: Optics, Focal surface detector, Trigger and Electronics System. An effective synergy between the parts constituting the instrument is of fundamental importance for achieving the EUSO scientific objectives. Optics, detector elements, system and trigger electronics have to be matched and interfaced coherently to obtain a correct response from the instrument. Scientific requirements have been of guidance for the conceptual design of the apparatus and in the choice among various possible technical solutions.

The observation from space calls for an approach different from that of the conventional ground based fluorescence experiments. For space application the instrument has to be compact as much as possible, highly efficient, and with a built-in modularity in its detection and electronics parts.

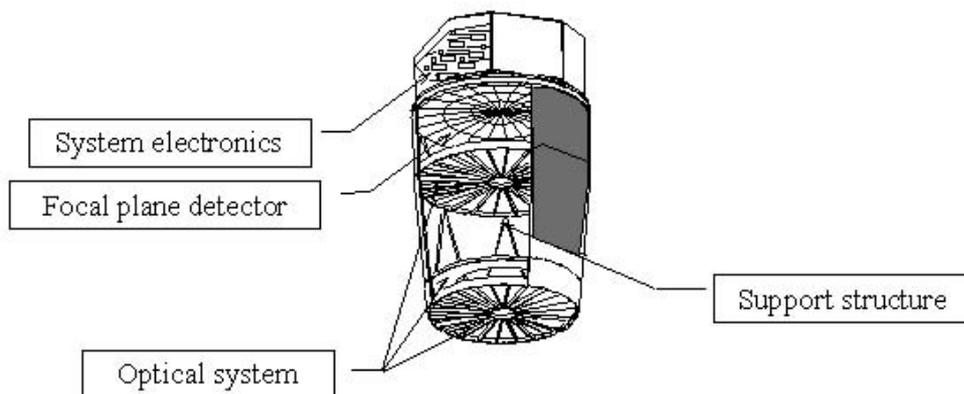


Figure 8 . View of the EUSO Main Telescope

1.3.1 The Optics

The optical system required for EUSO aims at finding the best compromise in the optical design, taking into account the suitability for space application in terms of weight, dimensions and resistance to the strains in launch and orbital conditions.

The optical system views a circle of radius ~220 km on the Earth and resolves 0.8×0.8 km² ground pixels: this determines the detector size to be adopted to observe the events.

The forgiving resolution requirements of EUSO suggest the consideration of unconventional solutions, identified in the Fresnel lens technology. Fresnel lenses provide large-aperture and wide-field with drastically reduced mass and absorption. The use of a broader range of optical materials (including lightweight polymers) is possible for reducing the overall weight.

The present Fresnel optical camera configuration study (FoV 60°) considers two plastic Fresnel lenses with diameter 2.5 m and iris diaphragm 2.0 m diameter.

1.1.2 The Focal Surface Detector

Due to the large FOV and large collecting area of the optics, the focal surface detector is constituted by several hundreds of thousands of active sensors ($\approx 2 \times 10^5$ pixels). The detector requirements of low power consumption, low weight, small dimension, fast response time, high quantum efficiency in UV wavelength (300–400 nm), single photoelectron sensitivity, limit the field of the possible choices to a very few devices. A suitable off-the-shelf device is the Multi-Anode Photomultiplier Hamamatsu R5900 series. These commercial photomultipliers meet closely the requirements imposed by the project. Pixel size, weight, fast time response and single photoelectron resolution are well adaptable to the EUSO focal surface detector. The organization in “macrocells” of the focal surface (a macrocell is a bi-dimensional array of $n \times n$ pixels) offers many advantages as easy planning and implementation, flexibility and redundancy. Moreover, modularity is ideal for space application. The Multi-Anode Photomultipliers represent, in this contest, a workable solution.

1.1.3 Trigger and Electronics System

Special attention has been given to the trigger scheme where the implementation of hardware/firmware special functions is foreseen.

The trigger module named OUST (On-board Unit System Trigger) has been studied to provide different levels of triggers such that the physics phenomena in terms of fast, normal and slow in time-scale events can be detected. Particular emphasis has been introduced in the possibility of triggering upward showers (emerging from the earth, “neutrino candidate”) by means of a dedicated trigger logic.

The FIRE (Fluorescence Image Read-out Electronics) system has been designed to obtain an effective reduction of channels and data to read-out, developing a method that reduces the number of the channels without penalizing the performance of the detection system. Rows wired-or and columns wired-or routing connections have been adopted inside every single “macrocell” ($n \times n$ pixels unit, ≈ 100 macrocells constitute the focal surface detector) for diminishing the number of channels to read-out.

4 EUSO Duty Cycle

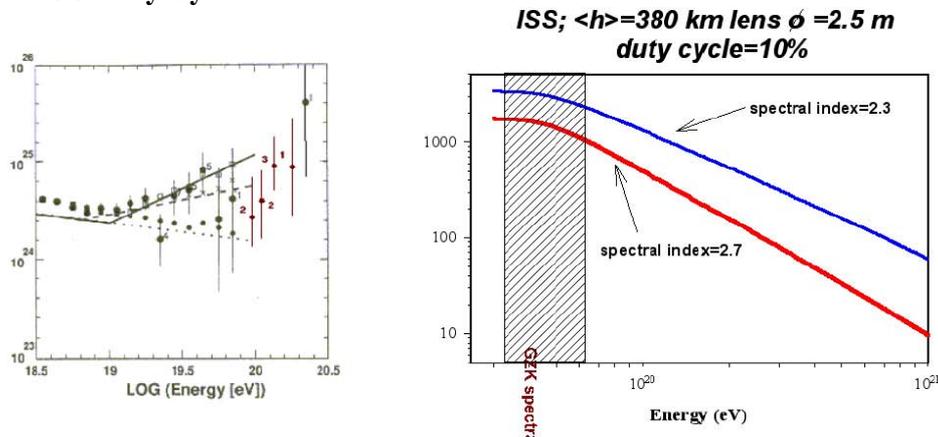


Figure 9. left)- This is Fig.10 from Yoshida et al., *AstroParticle Physics* 3, 1995, pp. 114: “Derived primary energy spectra expressed by Eqs. (11a) (solid line) and (11b) (dashed line) and the expected values in each bin simulated under the assumptions of these spectra with the energy resolution of the present experiment open squares and crosses). Black dots with error bars are the raw data. The case of a single power up to the highest energy is also shown by a dotted line and shaded circles.” The slope of the continuous line above 1019 eV is 2.3; for the dashed line the slope is 2.7. The superimposed red points are from Takeda et al., *Phys. Rev. Lett.* 1998, 81, pp.1163. **right)**- EUSO counting rates under the hypothesis of the two different spectral index assumed (see Fig.1 left).

The EUSO duty cycle has been estimated taking into account the following factors affecting the level of background:

The ISS night time; ground locations with significant light output, natural or anthropomorphic ; Lunar cycle ; Clouds in the FOV strongly affecting the detection or interpretation of the EAS ; ISS activities or contingencies that do not allow the operation of EUSO .

The likely EUSO duty cycle is resulting to be in the range 0.1-0.15; a more precise evaluation requires a detailed assessment of the various elements, in particular of the cloud related effects: we, in a conservative approach, use a value of 0.1 throughout this report.

5 Expected Results

The slope of the CR energy spectrum in the region of the GZK limit and above is poorly known because of the reduced statistics available (see Fig. 1, insert). The expected counting rates for energies e.g. above 10^{20} eV, are therefore difficult to define and are strongly dependent from the assumed extrapolation for the energy spectrum. As an example, we show Fig. 9 where the integral count rates are given respectively for a spectral index -2.7 and -2.3 : the counts per year above 10^{20} eV varies from 500 to 1300, accordingly. In the following, a spectral index of -2.7 is assumed (in a conservative way).

Figures 10 and 11 show the predicted number of EECR and neutrino events per year as a function of energy detected by EUSO in the original free-flyer and ISS configurations. Both configurations give comparable results within a small factor with the lower observational altitude of the ISS (380 km) counterbalancing somewhat the sensitivity afforded by the larger optics diameter of the free-flyer. The integral number of counts above an energy E for the two configurations is shown in Fig. 12 assuming the 2 year operational life of the free-flyer and a requested 3 year lifetime.

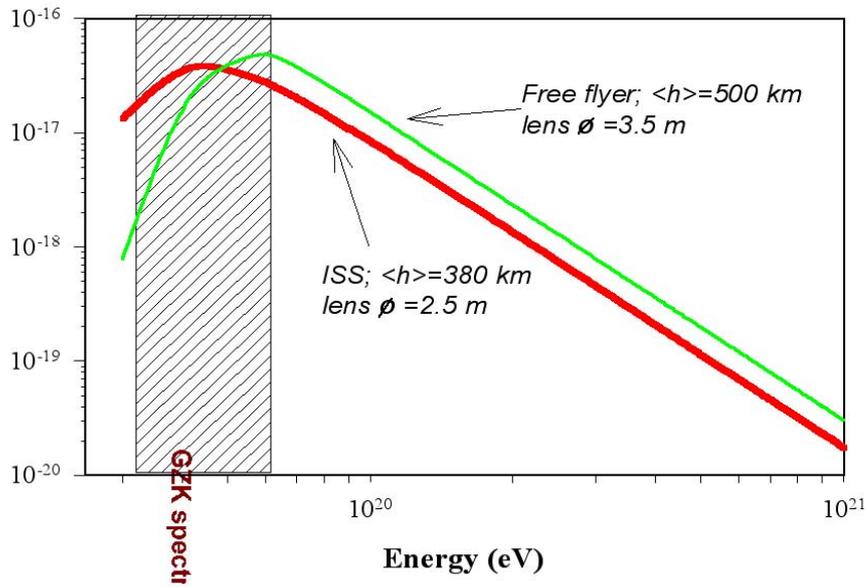


Figure 10. Differential EECR counting rate comparison between the ISS version of the EUSO and the original free flyer. The dashed zone shows the spectral region where structure induced by the GZK cutoff is expected. The lens diameter is the maximum external diameter allowed in each configuration.

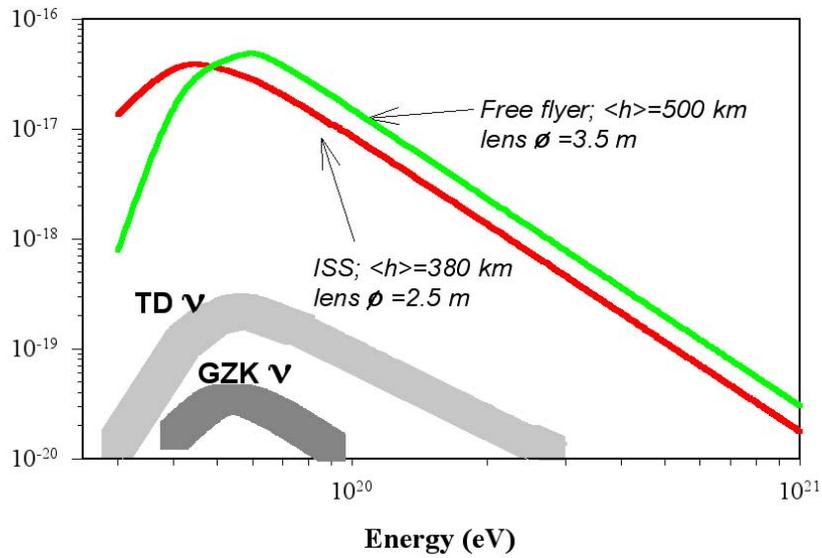


Figure 11. The differential flux of neutrinos predicted using the Topological Defects model of Sigl et al. (1998) and the GZK model of Stecker et al. (1991).

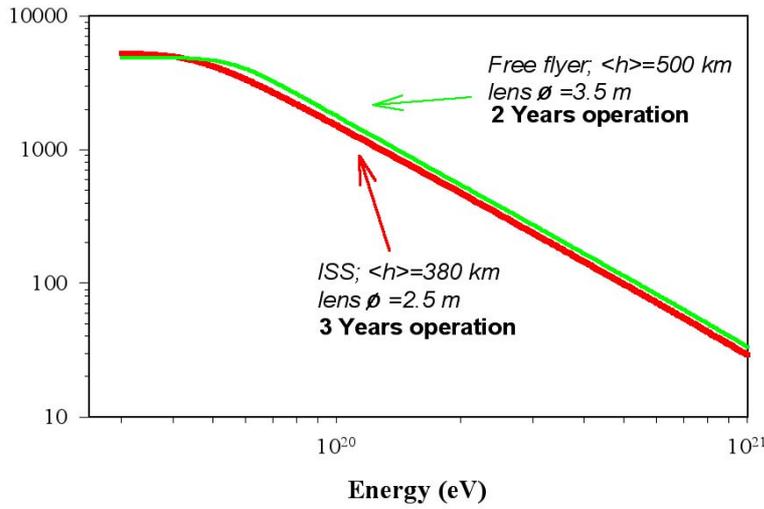


Figure 12. The integral count rates above an energy E predicted for the original free flyer proposal with 2 years of operations and the ISS configuration with 3 years operations.

5.1 Comparison with ground based observations

In the following, a spectral index of -2.7 is assumed whenever absolute values are quoted for the counting rates above 5×10^{19} eV. Currently only the AGASA and HiRes instruments are operational. The AGASA experiment near Tokyo, with coverage of about 100 km^2 (about $300 \text{ km}^2 \text{ sr}$), comprises a scintillator array for electromagnetic shower particles and an array of muon detectors. The HiRes experiment in Utah consists of two fluorescence detectors at a distance of 12.6 km and, in a first stage, just came online. Its aperture is energy dependent and rises from $340 \text{ km}^2 \text{ sr}$ at 10^{19} eV to $1000 \text{ km}^2 \text{ sr}$ at 10^{20} eV. The largest planned ground-base experiment is the Pierre Auger Observatory, presently under construction in Argentina. This will consist of an array of 1600 particle detectors covering 3000 km^2 and 4 fluorescence light detectors, similar to the ones used in the HIRES experiment. The hybrid detector allows cross calibration and a check of the systematic

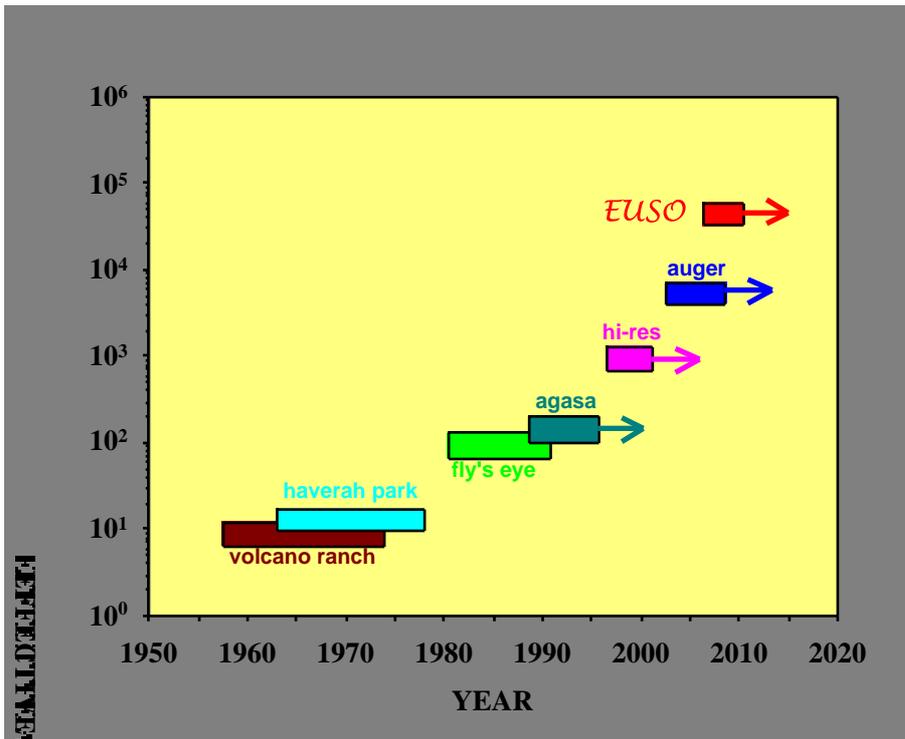


Figure 13. A Comparison of the EUSO effective area with ground based facilities

certainties inherent in each of the techniques. The construction is expected to be completed in 2004. By then the Auger observatory will have an aperture of $7000 \text{ km}^2 \text{ sr}$, leading to about 30 events per year with energies $>10^{20}$ eV. Though a second Auger observatory is planned in the northern hemisphere, it is not clear whether, and when, there will be funding for it. Auger will produce a comparable number of events to all previously

observed above 10^{20} eV in only 4 months. Nevertheless, a rate of 30 events per year is too small to follow the CR spectrum to such energies, or to obtain the detailed form of the spectrum with small statistical errors. At 10^{21} eV only about 5 events are expected in 10 years of operation. At least an order of magnitude more statistics is desirable. By the year 2006, provided Auger is completed on schedule, the world data set will comprise about 100 events above 10^{20} eV and perhaps one or two events above 10^{21} eV, if the spectrum continues without a GZK cutoff. The existence or non-existence of the GZK cutoff will most likely be established by then. However, a definite answer concerning the origin and possible identification of sources will certainly rely on the precise spectral form and on the arrival direction distribution. In both cases good statistics are vital to distinguish between different competing scenarios.

For EUSO an effective geometrical factor of 5×10^5 km² sr and an observing efficiency of 0.1 gives an expected event rate of ~500 per year for EECR with energies $\geq 10^{20}$ eV (no GZK suppression and spectral index of 2.7), or some hundreds per year with GZK suppression. The total number of events per year with the low-energy threshold of $>3 \times 10^{19}$ eV is >1700 . For EECR neutrinos the expected rates vary from a few per year (GZK processes, AGN, GRB sources, etc) to ~150 per year if top-down processes dominate.

6 Acknowledgments

A general acknowledgement is due to the EUSO Collaboration. I'm deeply indebted to M.C. Maccarone for her determinant contribution.

1. References
2. Scarsi, L., et al., "EUSO – Extreme Universe Space Observatory", Proposal for the ESA F2/F3 Mission, January 2000, and references therein (2000) <http://www.ifcai.pa.cnr.it/~EUSO/docs/EUSOproposal.pdf>
3. ESA/ESTEC, "EUSO Accommodation Study Report", ESA/MSM-GU/2000.462/AP/RDA, December 2000.

PHOTOMULTIPLERS FOR THE NEW TOF SYSTEM OF AMS-02 SPACE EXPERIMENT

Cristina Sbarra
for the AMS-Bologna collaboration
INFN, Bologna, Italy

ABSTRACT

The AMS-02 experiment is a cosmic ray spectrometer to be operated in space after 2003. The Time Of Flight system of the experiment has to withstand a huge magnetic field (up to 2.5kG), so the fine mesh photomultiplier Hamamatsu (R5946) were chosen for the TOF counters. Some studies were made on the tubes, and they are shown to be rather insensitive to the magnetic field up to about 4 kG, at angles between the field direction and tube axis up to about 30°. A comparison with a simulation of the fine mesh time response has also been made and the results seem quite encouraging.

1 Introduction

The *Alpha Magnetic Spectrometer* (AMS) installed on the International Space Station at the end of 2003, and will measure cosmic ray fluxes for at least three years in a low orbit (about 400 km) around the earth. After the successful operation of AMS-01 aboard of the Space Shuttle Discovery in a 10-days test flight carried out in June 1998, rigidity up to 1 TV, by using a superconducting

magnet which will provide a maximum field of about 0.8 T ¹. As a consequence, the TOF system with different values for the angle between the photomultiplier tubes and the field direction.

In the following a short description of the new TOF system is given and results are reported on the operation of the Hamamatsu R5946 photomultiplier tubes inside the magnetic field. The results of our first approach simulation are also reported.

2 The TOF system of AMS-02

The TOF system will provide the fast trigger of the experiment and the measurement of the time of flight of the particles traversing the detector with a resolution sufficient to distinguish upwards and downwards going particles at a level of at least 10^{-10} , and electron and antiproton up to about 2 GeV .

The AMS-02 superconducting magnet produces a field of about $2 \div 3 \text{ kG}$, of variable direction on the TOF planes and, thus, on the PMs. The field magnitude and the angle between the magnetic field and the tube axis are shown in figure 1 for all tubes of the system. After a market study, the Hamamatsu R5946 photomultiplier tube was considered the best choice and thoroughly tested for time resolution and pulse height response in magnetic field.

3 Experimental PM behavior in magnetic field

We placed the phototube inside an electro-magnet (maximum field 4kG) on a movable stand which could be rotated at a maximum angle of 90° with respect to the magnetic field. A red diode was used to make light on the PM. The charge signal from the photomultiplier was digitized by an ADC and registered by a PC-based data acquisition system. The photomultiplier tubes gain has been measured following the method suggested by ⁵). tests in magnetic field: one with high gain (PM R5946-WA9386), one with intermediate gain (PM R5946-WA9385) and one with low gain (PM R5946-WA9381). All three tubes were then operated at a gain of about 2×10^6 , with voltages of -1700V , -2000V and -2200V , respectively. The photomultiplier responses have been measured for various intensities of the magnetic field B and with different angles between the tube axis and the field direction, called θ in the following. Figure 2 shows the response of the tubes at different values of the magnetic field, normalized to that at $B = 0$, as a function of θ . For $\theta \leq 50^\circ$, the relative response is well above 50% up to the highest values of B . Only at $\theta = 60^\circ$ the response goes to zero for $B \geq 1000 \text{ G}$ ⁶).

¹The improved detector is nicknamed AMS-02 in the following.

4 Time resolution from data and Simulation of time response

The signal from the photomultiplier tube has been sent to a discriminator (with the threshold set at 30mV) and then to the stop of a 50ps/bin TDC, started by the pulse driving the LED. The resulting time distribution has a resolution of about 330ps, much worse than the 100ps expected for real scintillation light because of the rather longer duration of the LED light pulse. In fig. 3 are shown the mean transit times² and the time resolutions for tubes no. 9381 and 9386, as function of the magnetic field and for different values of θ (the times plotted are relative to the time at $B = 0$ of PM9381).

The most relevant observation is that the tube operated at higher voltage (no.9381) shows a shorter transit time and a better time resolution. The transit time generally gets worse with increasing angle for both PMTs, but it is more critical for the PM working at lower voltage (no.9386). This is also confirmed by our simulation, and gives us some hints on the possible way of arranging the PMTs in the new TOF system of AMS-02. Moreover a strong correlation is observed between time resolution and transit time for both tubes.

response in magnetic field by solving the photoelectron first order differential equation of motion (in the velocity and position unknowns). We used the Runge-Kutta numerical approximate solution in finite time intervals⁷). We started with many photoelectrons from the fine mesh photocathode, and followed them up to the first dynode, considering the electric and magnetic field strengths and the angle between them. To simulate the fine mesh time response, we decided to extract randomly only one secondary electron at each dynode, up to the final anode. The photoelectron extracted was given an angular distribution generally flat over the downwards hemisphere (i.e. towards the anode). The only case in which such secondary electron kept the ingoing direction, was when it went elastically scattered. We followed the energy distribution of the secondary electrons emitted (SEE), whose general shape is given also in⁸). Finally, we chose one of the secondary electrons emitted at each dynode, and cumulated the time it took to get to the next one, up to the anode. At the end, we got a distribution of time of arrivals at the anode as a function of θ and magnetic field strength, for different simulated HV. Figure 4 shows our simulation of the fine mesh time response and the experimental results as function of the angle between \mathbf{E} and \mathbf{B} and at different magnetic field strengths for PM no.9386 (HV= 1700) and for the PM no.9381 (HV= 2200 V). The time plotted is relative to the time at $B = 0$ of the low gain PM. The worsening of the transit time with respect to the angle θ , both experimental than simulated, is most critical for the highest gain fine mesh PM (no.9386). This is true also for the dispersion of the measured time dispersions⁶), see figure 3.

²transit time: time delay with respect to the LED pulse.

5 Conclusions

A serious worsening of the TOF counters time response occurs when their phototubes operate at angles above 30° , nearly independently of the applied magnetic field. In order to reduce as much as possible the time resolution degradation, we had to constrain the mechanical design of the TOF so to avoid the larger angles between tube axis and field direction (θ greater than 30°). Anyway, the data of the tests made combined with the fine mesh simulation results, suggest us also to higher the HV for the critical cases in which the PM must yield an angle $\theta > 30^\circ$.

References

1. S.P. Ahlen *et al.*, Nuclear Instruments and Methods **A 350** (1994) 351.
and
2. J. Alcaraz et al., Phys. Lett. B, 484, p.10, 2000, 472, p.215, 2000J
3. D. Alvisi et al., Nucl.Instrum.Meth. A437 (1999) 212.
4. W.R. Leo, "Techniques for Nuclear and Particle Physics Experiments", Springer-Verlag (1987).
5. B. Bencheick et al., Nucl. Instrum. and Methods A315(1992) 349.
6. L.Patuelli, "Studio delle prestazioni in campo magnetico dei fototubi 'Fine-Mesh' per il sistema T.O.F. dell'esperimento spaziale AMS02" Laurea Tesi, Bologna (2000).
7. W. Hpress et al., "numerical Recipes in Fortran 77: The Art of Computing", Cambridge University Press, also available at the url: <http://www.nr.com/>.
8. G.Barbiellini *et al.*, "A simulation study of the behaviour of fine mesh photomultipliers in magnetic field " **A 362** (1995) 245-252.

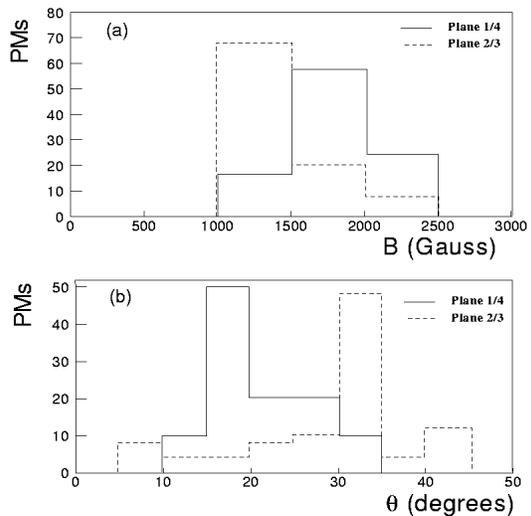


Figure 1: Magnetic field intensity (a) and angle (b) distribution for PMTs of plane 1 (solid line) and 4 (dashed line).

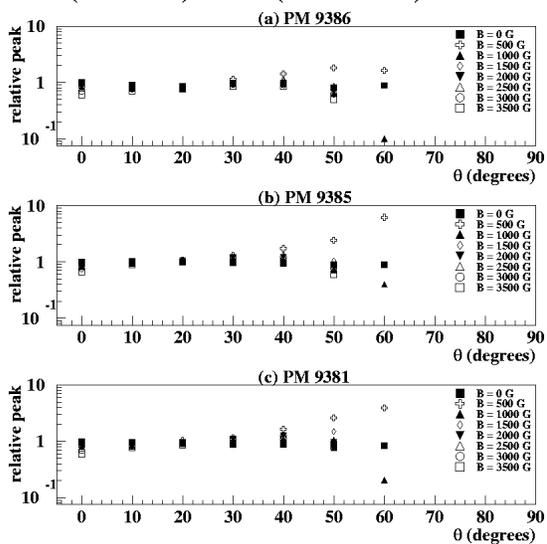


Figure 2: PMTs behavior versus θ angle. between the tube axis and the field direction, as measured for at different values of the magnetic field B (normalized to $B = 0$).

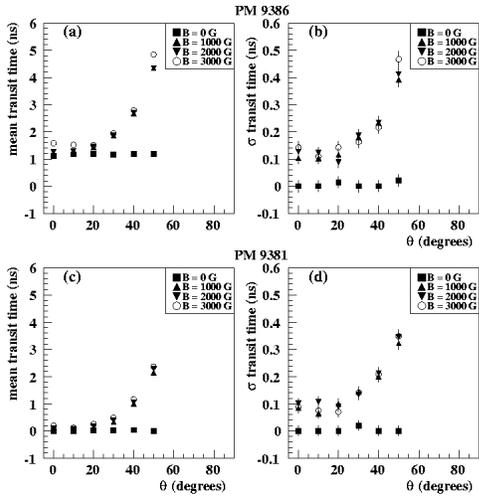


Figure 3: Mean transit time and time resolution as function of θ and for different values of B , for PM no. 9386 (a,b) (HV=1700V) and PM n.9381 (c,d) (HV=2200V). For both PM the time at $B = 0$ of PM9381 is subtracted. The PM operated at higher voltage shows the shorter transit time and the best time resolution.

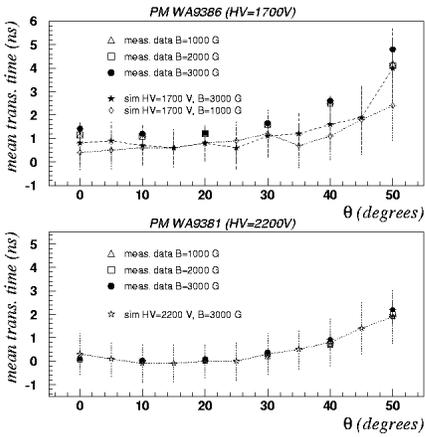


Figure 4: Measured and simulated mean transit time for the PM9386 (HV=1700) and for the PM9381 (HV=2200) at various θ and fields (the time at $B = 0$ of low gain PM is subtracted). The PM working at higher voltage shows a shorter transit time also in such simulation.

The gamma-ray side of the Universe

P.Fleury	Gamma Astronomy - Results from Ground Observations
G.Barbiellini	Science and Technology: The AGILE Scientific Instrument
M.Tavani	Gamma -ray astrophysics and the AGILE mission
R.Bellazzini	The GLAST Gamma ray large area space telescope
A.Morselli	Gamma ray astroparticle physics with GLAST
R.Battiston	AMS- γ : High Energy Photon Detection with the ALPHA Magnetic Spectrometer on the ISS
A.Celotti	The physics of Active Galactic Nuclei
A. De Angelis	The Simulation of Astroparticle Experiments and the GEANT4 Toolkit
G.Kanbach	MEGA - a new telescope for medium energy gamma-ray astronomy
S. Rebecchi	An introduction to ASDC a new ASI facility
V. Cocco	Geant Simulation of the AGILE Gamma-Ray Imaging Detector
C. Pittori	Presentation of the AGILE Recontruction Method and Kalman Filter Algorithms for Gamma-Ray Silicon Detectors in Space

Frascati Physics Series Vol. nnn (2001), pp. 000-000
TITLE CONFERENCE - Conference Town, Date dd mm, 2001

GAMMA ASTRONOMY - RESULTS FROM GROUND OBSERVATIONS

Patrick FLEURY
IN2P3 - Ecole Polytechnique, 91128 Palaiseau (Fr)

ABSTRACT

The TeV energy domain is covered by the use of ground based detectors. This indirect technique makes use of the atmosphere as the sensitive volume, thus allowing large detection areas. It has become a mature technique along the last decade, still open to significant development avenues. Results to be presented here concern a few objects from a variety of astronomical entities, belonging to the Galaxy such as plerions and supernova remnants, and extragalactic ones, i.e. blazars. The evidence of activity up to TeV energies brings in important clues to the study of these objects, and the very large detection area also favours variability studies. The coming generation of detectors will reach sources of lower luminosities and their sensitivity will be extended to lower energies, down to few tens of GeV. The common exploration with GLAST of the essentially untackled 1-100 GeV energy domain is very promising, the highest domain for which the Universe remains transparent up to cosmological distances.

1 Introduction

Speaking of ground based observations at a school focused primarily on space Astronomy, it is necessary to specify their respective domains of energy. Let's simply refer to subdivision first suggested by T.Weekes ¹⁾ and revised by Hoffman et al. ²⁾ (see also Ong ³⁾) between High-Energy and Very-High-Energy - HE / VHE - respectively from 30 MeV to 30 GeV and up to 30 TeV. The frontier is representative of the relevant techniques, direct detection from space for HE (COSB, AGILE, GLAST ...) and ground based air showers Cerenkov detectors for VHE. Up to now, there is no frontier contests as there exists an unexplored region extending over about a decade on both sides, from 3 to 30 GeV and from 30 to 300 GeV, with essentially no data. Needless to say that many open questions bump against this region so that its exploration does constitute a major objective.

2 The Atmospheric Cerenkov Technique

The ground based observations with the use of Atmospheric Čerenkov Telescopes (ACT) allow sensitive areas several order of magnitudes larger than the direct detection from space, typically 10^5m^2 against 1m^2 . Besides, the atmosphere constitutes a very massive 'calorimeter', any estimator of the shower amplitude is an estimator of the primary energy. On the other hand, the ground based observations suffer from the difficulty to separate γ -rays from charged cosmic rays (CRs), since there is no veto against charged particles as can be done around a compact detector in space. The good angular resolution - of typically $\leq 0.1^\circ$ on each event - favours the search of point sources out of an intense background of diffuse CRs. However, the success of the Čerenkov techniques relies strongly on the demonstrated possibility to separate γ -events over hadron events.

Its foundation dates back to the Galbraith and Jelley experiment in 1953 ⁴⁾ which established that the Čerenkov light first observed in liquids can be also be generated in gases. The detection of air showers through the Čerenkov light has been pursued ever since. The indirect detection of the charged secondaries by Čerenkov light gives access to much lower γ -energies: the light propagates with limited absorption down to the ground while charged secondaries reinteract and get absorbed. This light collection is handled by

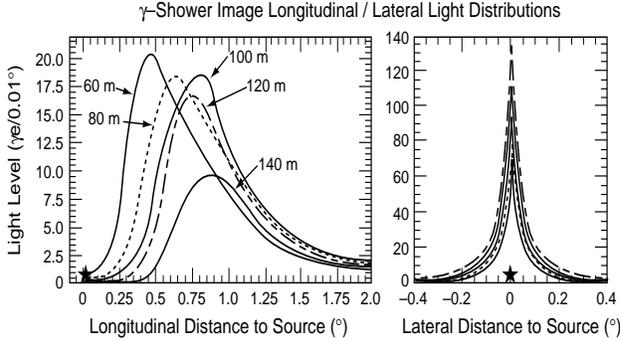


Figure 1: Image of γ initiated air showers. Left: longitudinal and transverse shower profiles from simulations of TeV γ events. Right: an event observed with the ≈ 600 pixel camera of the CAT telescope.

focusing devices. Fast photomultipliers (PMT) can be exploited allowing to beat the night light, which amounts to about 1 photon/ns m^2 sr mostly from the Milky Way. The prices to pay are that of low duty cycles (moonless and cloudless nights) and small angular acceptances. Some attempts to turn this difficulty around are presently in a validation phase : running at higher altitudes to reach moderate energies, they are also based on improved detection techniques, using water detectors for MILAGRO, near Los-Alamos or "resistive plate chambers" for ARGO in Tibet ^{6) 5)}.

2.1 The imaging

The 'ground braking' experiment is that of the Whipple observatory, nearly 20 years after the large dish telescope of 10 m diameter has seen its first light (in 1969). The break-through came from the imaging of the shower based on a multi-pixel camera, with 37 PMTs at first. The images happen to fulfill two independent roles, their position in the camera plane provides an information on the direction of the shower, and their shapes allow a differentiation of gamma (γ) versus hadron showers. The γ induced shower propagates grossly from 12 km to 3 km asl (above sea level) with a lateral expansion of ≈ 20 m from axis. The shower propagation can be described as a propagation at light speed of a narrow and thin bright disc illuminating downwards, within a limited angular

cone. The image of this thin bright rod, from a point of observation within ≈ 100 m from the shower axis, is compacted into an elliptic shape aligned with the source-point, at the center of the camera (as the telescope is tracking that source). CAT shower images shown in Fig1 illustrate the imaging concept. The 600 PMT camera affords an image resolution of about 0.1 degree.

The shower image is analysed in terms of its first order moments, according to M.Hillas prescriptions upgraded by M.Punch's "supercut" method ⁷⁾. The relevant information is the total number of photons, the main axis position, the length and width. Conventionally the signature of a source signal is expressed by the statistical significance of the peak at zero of the parameter α which measures the misalignment of the image main axis from its expected radial disposition. The CR events should be uniformly distributed in α .

Thanks to the high resolution camera, higher order information can be exploited in the case of CAT. Each image is fitted in direction, energy, distance of impact and source point, against templates which have been simulated according to sets of discrete values of these parameters (see Fig.1, left frames). Each individual γ shower image can then be reduced in terms of a two coordinate point in the sky map, plus an energy estimator ⁸⁾.

A few observatories established since 1990, after the first success of the Whipple 10m diameter telescope, have reached comparable sensitivity with smaller collection mirrors, either exploiting a stereoscopic multi-telescope techniques in the case of HEGRA or a higher resolution imaging in the case of CAT and CANGAROO. Both lines of improvements have proven to be very useful, and both will be invested jointly in the future programs of HESS, VERITAS and CANGAROO-II,III, while for MAGIC, which places the emphasis on highly improved imaging, the addition of a second telescope is also considered (see Table 1).

2.2 Distributed sampling

The imaging technique is not the only method. Large arrays of moderate size telescopes, each one equipped with a unique PMT have been exploited with some success. In fact, the first confirmation of TeV emission from the Crab nebula observed by Whipple came from the Themistocle array in the French Pyrénées consisting of 18 telescopes of only 0.2 m² each.

To exploit the 1-2 ns duration of the Čerenkov pulse, the difference in

Imagers Site	Countries	Nb of telescopes x Area	Nb of pixels	E_{threshold} (GeV)
Whipple <i>North</i> Arizona	USA UK, Ireland	1 x 75 m ²	91 (+18) 380 (+111)	250 100-200
7 TA <i>North</i> Utah	Fapan UK, Ireland	7 x 3 m ²	3 x 256	250 100-200
Cangaroo <i>South</i> Woomera Australia	Japan Australia	1 x 30 m ² 1 (→4) x 57 m ²	512 552	300
HEGRA <i>North</i> Canaries	Germany Spain	5 x 8.5 m ²	271	1000 (1 t.) 500 (4 t.)
NARRABRI <i>Sth</i> Narrabri Australia	UK Australia	[3 x 42m ²] same mount	91	300
CAT <i>North</i> Thémis France	France Tch. Rep.	1 x 18 m ²	546 (+50)	250
TACTIC <i>North</i> Mt Abu, India	India	1 x 18 m ² +2 no-image	349	300
Solar Plants Site		Nb of Heliostats x Area		
CELESTE <i>North</i> Thémis France	France Tch. Rep.	40 x 54 m ² → 53 x 54 m ²	one per heliostat	35 (trigg.) 60 (anal.)
STACEE <i>North</i> Sandia Arizona	USA Canada	48 x 37 m ²	id.	100
Keck-Solar2 <i>North</i> Barstow California	USA	32 x 40 m ² → 64 x 40 m ²	id.	tests in progress
Other ACT				
PACT <i>North</i> Pashmari India	India	25 x 105m ² 25 x 105m ²	7 mirrors /telescope	3000
Non ACT Site		Detector Area	Altitude	
Milagro <i>North</i> Fenton Hill, NM	USA	Water Čerenkov m ²	2.6 km	500
Tibet HD <i>North</i>	Japan China	Scintillators	4.3 km	3000
ARGO <i>North</i> Tibet	Italy China	RPC detect. 50 → 10 ⁴ m ²	4.3 km	tests in progress

Table 1: Existing ACT's. *In this table North and South refer to the northern and southern hemispheres*

Source	Type	Red-shift z	Date of discovery	EGRET	Grade (T.W.)
Galactic sources					
Crab Nebula	Plerion		1989	yes	A
PSR 1706 - 44	Plerion ?		1995	no	A
Vela	Plerion ?		1997	no	B
SN 1006	SNR		1997	no	B
RXJ 1713-3946	SNR		1999	no	B
Cassiopeia A	SNR		1999	no	(C) A
Centaurus X3	Binary		1999	yes	C
Extragalactic					
Markarian 421	XBL	0.031	1992	yes	A
Markarian 501	XBL	0.034	1995	yes	A
1ES 2344 +514	XBL	0.044	1997	no	C
PKS 2155 6304	XBL	0.116	1999	yes	B
1ES 1959 +650	XBL	0.048	1999	yes	B
1ES 1426 -427	XBL	0.129	2001	no	A

Table 2: *Source Catalog*

trajectory of light must be compensated as this extends over a millisecond scale between telescopes. This timing correction must be dynamically built-in, since the time delays vary continuously during the tracking of a source. The foot print of γ -events will then be made of the response of each telescope in number and in arrival time of the collected photons. This "light pool sampling" pertains information on the shower core, on its axis. Its spatial distribution - from telescope to telescope - plays a role similar to that of the image configuration to separate hadron from γ -showers.

The sampling method is used by experiments presently exploiting solar plant arrays, CELESTE, STACEE and SOLAR-2. Collected by large "heliostats" of 30-50 m² the light is focused at the top of a central tower. Following a scheme first proposed by T.Tumer⁹⁾, a secondary optical device is set at the tower top which distributes the collected photons to individual PMTs, one per heliostat. This restores the independent collection of photons for each heliostat, a prerequisite to correct for the time lags which vary continuously. The huge total mirror area permits to lower the threshold to a few tens of GeV.

Experiment Site	Countries	Focal l. / Diameter Nb tel. x Mirr. area	Nb of pixels	year of first light
Cangaroo-III <i>South</i> Woomera Australia	Japan Australia	F/D = 8/10 (m) 4 x 57 m ²	512	2001 (2003)
HESS <i>South</i> Namibia	Germany France, UK	F/D = 15/12 (m) 4 (16) x 100 m ²	800	2002 (2004)
VERITAS <i>North</i> Arizona	USA UK, Ireland	F/D = 12/10 (m) 7 x 75 m ²	499	2005
MAGIC <i>North</i> Canaries	Germany Spain, Italy	F/D = 20/17 (m) 1 (→ 2) x 220 m ²	800	2002
MACE <i>North</i> Canaries	India	F/D = 20/17 (m) (→ 2) x 220 m ²	800	2007

Table 3: Future ACTs

3 Early motivations and new avenues of high energy gamma astronomy

For several decades, the unique evidence of very high energy phenomena was that of cosmic rays, a puzzling phenomenon by its huge amplitude, of about 1% of the mass energy of the Galaxy. The series of space detectors, from SAS-2, to COBE and EGRET aboard CGRO, were aimed primarily at the search of CR acceleration sites and the determination of acceleration mechanisms. Before EGRET, there were no other known HE phenomena, apart from a few pulsars and the serendipitous discovery by COBE of a single extragalactic source, namely 3C-273 already known as an active galactic nucleus (AGN). In the eighties, at multi TeV energies, violent activities from a few sources (Cygnus X3, Hercules X1) had been claimed by air shower experiments either based on charged secondaries or Čerenkov detection. Rare are the authors still giving much credit to these observations ... which did contribute significantly to the deployment of present VHE activities. A new era started with the last decade, with the launch of CGRO and with the first firmly established evidences of detection of TeV γ -rays from the Crab nebula by the Whipple group¹²⁾ in 1989. The main contributions of EGRET, the HE gamma detector aboard CGRO, are the discovery of the emission from ≈ 70 blazars. Other topics concerning the Galaxy have been approached, mostly :

- the observations of a few supernova remnants, possible sites of CR ac-

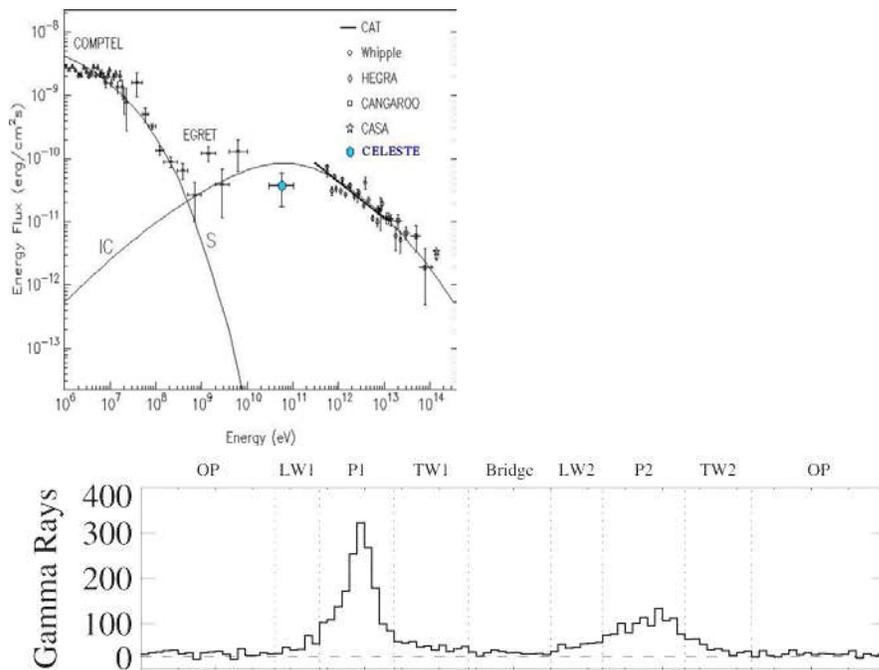


Figure 2: *Top, the two component Crab nebula (unpulsed) emission: the (S) Synchrotron and (IC) Inverse Compton. The behaviour in the intermediate energy region needs to be confirmed; in particular, EGRET data are extracted from the above phasogram dominated by the pulsed contribution.*

celerations;

- the energy spectrum of the energetic pulsars;
- the improved cartography of diffuse Galactic γ -rays;
- an abundant wealth of unidentified sources, partly located in the neighbouring molecular clouds which constitute the Gould belt ¹⁰). The VHE results from ground based observatories which we shall now concentrate upon, are a way complementary to those from HE as they deal with the same sort of objects, the pulsars and their nebulae, the supernova remnants and the blazars.

4 The Crab plerion : a supernova remnant nebula energized by a pulsar

The Crab nebula originated in a supernova noted by the Chinese astronomers about 1000 years ago. Seen at all wavelengths, from radio to TeV energies, it is energized by the collapsed residual star, a 33 ms pulsar which is itself observed from radio up to GeV energies. Such a system is designated as a plerion. The pulsed emission is of course attributed to the pulsar while, rather conventionally, the whole unpulsed emission is attributed to the nebula.

The signal obtained from ground observation around 1 TeV, was much too intense to match in continuity with either the pulsed or unpulsed lower energy emissions. For this indirect and by then a new technique, the question of flux calibration had to be settled. This was obtained with the convergence of data from the Themistocle experiment which extended the observations to over 10 TeV. So was clearly established by ACT, that a new process was at stake, able to generate the most energetic γ -rays ever detected.

As no sign of pulsation could be evidenced, the process was assigned to the nebula. This was not quite unexpected and a scheme was at disposal, at least since 1965, when R. Gould ¹³⁾ published his paper with the explicit title : "High Energy photons from the Compton-Synchrotron process in the Crab nebula". Energetic electrons in a magnetized environment, thus producing photons by synchrotron emission, may boost some of these photons up to TeV energies by Compton scattering. This is referred to as the "inverse Compton", because the observable signal can arise only from those photons recalling backward, in the direction of the incoming electrons and borrowing a large part of its energy. These high energy γ rays provide us with a rather conform image of the electron spectrum, somewhat distorted however at the highest energies by the Klein Nishima relativistic regime. The spectral energy distribution (νf_ν) shown in Fig.2 illustrates this two regime emission of the Crab nebula.

Soon after this observational break-through, a detailed model of the Crab nebula emission has been worked out by A. Harding and O. DeJager ¹⁴⁾ based on a nearly parameter free MHD model of the nebula. A recent version of this model has been discussed by O.DeJager ¹⁶⁾. Other descriptions have been proposed, but the basic scheme of R. Gould has been preserved, as worked out in the phenomenological analysis by Hillas ¹⁵⁾, leading to the evaluation of magnetic field and that of the electron spectrum within the nebula.

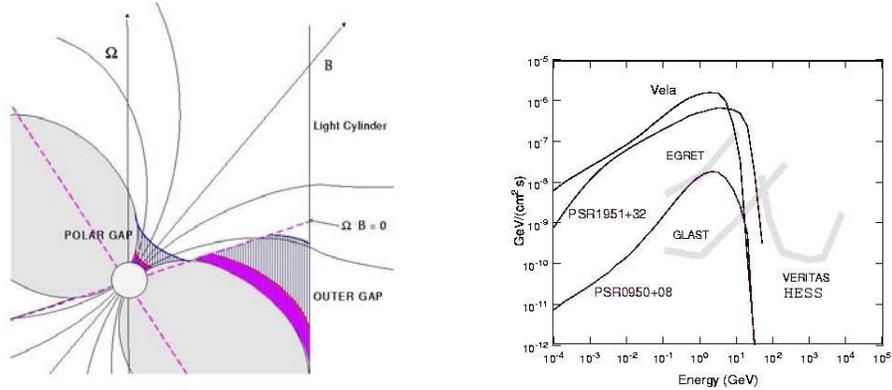


Figure 3: *Right : the polar cap and the outer gap models as visualised by A.Harding. Left :three pulsar candidates for high energy cutoff, with indications of the domain of detector sensitivity (also from A.Harding).*

As we shall see later, the blazars present convincing evidences for a similar two level process, of synchrotron and inverse Compton emissions referred as the SSC mechanism, for Self Synchrotron & Compton. In this perspective, the Crab nebula appears as a superb laboratory for high energy phenomena occurring in other contexts within the cosmos. The SSC scheme for the Crab nebula has got quite a large consensus although the relevant data remain scarce. There is a wide energy region ≈ 3 to 300 GeV, void of any firm observation. The unpulsed emission is well observed in X-rays by the GRIS balloon flights and it extends in the γ MeV domain as seen by COMPTEL (aboard C-GRO). But the pulsed regime largely takes over in the EGRET energy domain and it becomes even questionable to sort out any isotropic continuum from the pulsar phasograms, see Fig.2-bottom. However, as seen on the EGRET Crab phasogram from Fierro et al. ¹¹⁾, in Fig.2-down, for all data above 100 MeV, any continuous contribution appears as marginally small. Beyond the unobserved energy domain, the situation has again reversed, with a totally unpulsed contribution. The cumulated spectrum from all ACT observatories displayed a smooth energy decrease, well fitted by a unique power law. More recent data do show some evidence of a curvature compatible with a turn over around 100 GeV. The contribution from the CELESTE experiment, at about 60 GeV, brings in a first direct confirmation of the SSC prediction.

5 Pulsars

Pulsars are well defined objects, with a mass of ≈ 1.4 solar mass within a radius of some 10 km (resulting in a same momentum of inertia for all). At birth, they inherit of the star rotation momentum and of its compacted magnetic field. As the magnetic dipole and the rotation axes are usually not aligned, the sweep of the magnetic field generates a powerful electric field pulling electrons out of the star surface, and accelerating them. The pulsed radiation is emitted by the electrons from within the rotating magnetosphere, by bremsstrahlung, synchrotron, or curvature radiation, or by Compton interaction (see Fig.3).

Although there is no conceptual uncertainty about the overall machinery, the debate is still open on whether the radiation arises from near the star surface in the magnetic pole region or from some distance away within the magnetosphere. These are respectively the "polar cap" and the "outer gap" models (Fig.3-left).

As is often the case, the highest observable energies constitute a criteria to distinguish between the two classes of models. The polar cap predicts a sharp cut-off in energy while the outer gap leaves a priori more flexibility. Seven pulsars only (out of a few hundreds) are emitters up to the GeV domain as observed by EGRET. Their cut-off energies fall into the no man's land of few tens of GeV and will be measured with GLAST (Fig.3 right). But ground based observatories could do the job before, either with HESS observing PSR1706-44 from the southern hemisphere, and with MAGIC or even sooner with CELESTE or STACEE observing PSR1951+32 from the northern hemisphere which have both an expected cutoff at ≈ 40 GeV ¹⁶).

6 Search for cosmic ray acceleration sites

As recalled above, the CRs constitute the most striking evidence of non thermal very energetic phenomena in the Galaxy. A large effort has been focused on this issue which indeed has remained the first motive for the CGRO program. For the last few decades, the guide line has been to investigate primarily the regions of shocks at the outskirts of supernova remnants colliding against neighbouring dense molecular clouds. Shell type nebulae present X-ray evidences of synchrotron emission from energetic electrons. Whether protons are also accelerated - which could then feed the Galaxy with CR's - is the question.

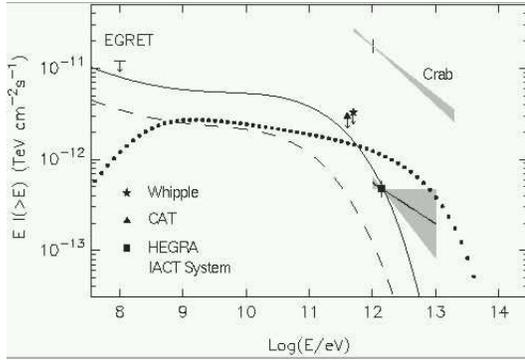


Figure 4: *The Cassiopeia A signal is now firmly established on the basis of 232 hours of cumulated observation by HEGRA. The solid and dashed lines represent evaluations of Inverse Compton productions while the dotted curve is the expectation for a CR acceleration site (γ 's from π^0 decays) .*

If so, some of them interacting with nearby matter, will produce π^0 s immediately decaying into γ -rays. An E^{-2} power law, as expected from shock wave acceleration, which is also needed to account for the $E^{-2.7}$ CR distribution (in as much as the $E^{0.7}$ law still holds for the leakage out of the Galaxy instead of $E^{0.3}$ as suggested by the CR modelisation of Strong et al. ¹⁷). This E^{-2} fall off matches with the sensitivity scales of the space and ground detectors, so that a genuine CR source could normally be tracked down from GeV to TeV energies. Nevertheless, out of 20 candidate sources, EGRET has detected energetic γ -rays from the few following objects : W44, W51, W63, Tycho, γ -Cygni, and IC448, none of which could yet be tracked down by ACT observatories.

ACTs have made at least as many attempts to reach altogether about a similar score, with positive observations from even fewer objects (see Table 3). The HEGRA team invested a huge effort - with 232 hours on source - to track down γ emission from Cassiopeia-A although this SNR has not been seen by EGRET (see Fig.4). ¹⁸ Altogether, this large effort has led to a rather puzzling situation which differs severely from the expected one-to-one correspondence between GeV and TeV observations. Whether more exotic hypotheses - acceleration at the outskirts of large bubbles or by Galactic γ -ray bursts ...- must or not replace the twenty years old guess in favour of SNRs is a licit question.

7 Quasars and Blazars

Alike TeV γ -rays from the Crab nebula, the first extragalactic TeV γ -rays were detected by the Whipple group. The observed source Markarian-421 (Mrk-421) was published in 1992 ⁷⁾. Its more than 5σ statistical evidence resulted from the reprocessing of the data with an improved version of the Hillas momentum analysis, the so-called "super-cut" method developed by Michael Punch in the course of his doctoral thesis. Four years later, another extragalactic source was discovered, again by Whipple, Markarian-501 (Mrk-501), an object similar to Mrk-421. These observations have been confirmed by other groups (such as HEGRA, CANGAROO and CAT).

7.1 What is a Blazar ?

At the time of these discoveries neither of these two objects had been seen by EGRET, as if there was a discontinuity between the HE and VHE domains as was indeed the case for Crab pulsar & nebula). Of the extragalactic EGRET sources that could be matched with an object previously seen at some other wavelengths all but one, Cen-A (a nearby galaxy) were blazars. Now understood as AGN having an associated jet in our direction, they are so named after the BL-Lacerta galaxy. The relativistic Doppler boost $\delta = [\Gamma(1 - \beta\cos\theta)]^{-1}$, with a typical Lorentz factor $\Gamma \approx 10$ enhances the observed energies in proportion of δ ($\delta \approx 2\Gamma$ at $\theta \approx O^\circ$), while the time scale is reduced as $1/\delta$. The apparent flux is strongly enhanced, according to $L_{app} \approx \delta^4 L$. Alike the Crab nebula, the blazar emissions extend over a very large domain of energy, presenting a broad double structure with an energy separation by factor of $\approx 10^8$ between both maxima in the spectral energy distribution (νf_ν). But, from blazar to blazar, there are wide variations of the overall scale: the first bump may lie in radio or up to X-ray wavelengths, the second bump being shifted accordingly. They are designated either as radio-blazar and X-ray-blazar (RBL / XBL) or, equivalently, as low and high energy blazars (LBL / HBL). The first bump is commonly attributed to synchrotron radiation of energetic electrons injected in a magnetized plasma (with the expected polarization). The high energy bump has often been tentatively attributed to accelerated protons with subsequent π^0 productions (decaying into 2γ s). Such a scheme would provide a direct relation between VHE γ -rays and VHE neutrinos (decay products from

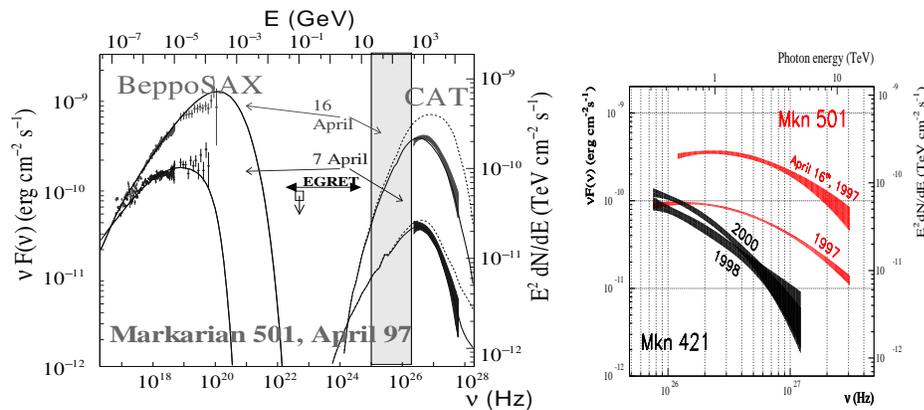


Figure 5: Spectral energy distributions ($\nu f_\nu = E^2 dN/dE$) for Mrk 501 during the flare of April 1997 (right) from X rays to TeV domains. Details of the TeV part for Mrk501 and for Mrk421 (left).

$\pi^{+/-}$). The alternative is a purely electromagnetic process as for the Crab nebula.

In Fig.5, the flare of Mrk501 in April 1997, illustrated with the CAT data ¹⁹⁾, together with the X-ray observations by Beppo-Sax satellite. Although there exists no data in between the keV and TeV regions, the double bump structure is at least not contradicted by the data which approach the respective maxim, from below in the X-ray domain and from above in the TeV domain.

7.2 Blazar variability

Blazars display irregular behaviour, varying by large factors in amplitude on small time scales, typically by factors of 10 and duration of as short as a day. Ground observations, at VHE energies, fully confirm this behaviour and even reinforce it significantly. The two best observed Markarian objects often lie below the present sensitivities and can raise up to 10 times the level of the Crab. The light curve of Mrk-501 for its flaring period of March 1997 has occasionally been quoted ²⁰⁾ as the founding picture of the ACT astronomy by proving the agreement and complementarity of five different observatories.

The evidence of a time correlation between X-ray and TeV emissions is

illustrated in Fig.5. The short time variability pleads for a compact region of emission logically located at the onset of the jet structure. The time coincidence of X-ray and γ -ray flares pleads for a single primary excitation within a same confined space. Variability on a time scale much shorter than a day have been observed on several occasions. The fastest change of rate ever observed is probably that of Mrk-421 in 1996 May flare, with a rate doubling in 10 minutes of time. This is about the light transit time from the Sun to Earth. Even corrected for a time Lorentz compaction of typically a factor of 10, this remains a surprisingly small region of emission. This fast time variation constitutes a key element in the elaboration of a model of emission.

The ACTs give access to short time scales and this will remain a superiority of ground based observatories. While space detectors are better suited for large surveys, the variability is best measured - on a given source - by the ground based detectors, in proportion of the geometrical collection area.

The presently unobservable energy domain should become soon the best measured domain as it will be accessed by the ACTs and by GLAST. Cross checks will become possible within this highly strategic energy domain and, complementary information will be supplied. At full extensions of their foreseen arrays VERITAS and HESS will reach a sensitivity to variability 10^5 higher than that of GLAST. The complementarity will remain essential.

7.3 The SSC - and EC -models

The fast variability is in favour of the primary electrons instead of protons, as the synchrotron radiation is an efficient mechanism to damp swiftly the electron energies. The target photons for the inverse Compton interactions may proceed also from other sources than the synchrotron effect, e.g. the thermal emission from the accretion disk. This is rather likely for XBL blazars which present emission lines. This extension of the SSC model is the EC model (for External Compton). Of course, too much damping will prevent electrons to reach the highest possible energies. An anti-correlation between intensity and energy is thus expected which seems to hold pretty well, at least statistically. As illustrated by Fig.6, ≈ 70 blazars are subdivided in 5 sub-samples²²⁾ according to their absolute intensity (corrected for distance) at radio frequency of 10 GHz. All five classes show the double bump typical of the SSC or EC scheme. The intrinsically brightest blazars are low energy (LBL) type. Conversely, the HBL

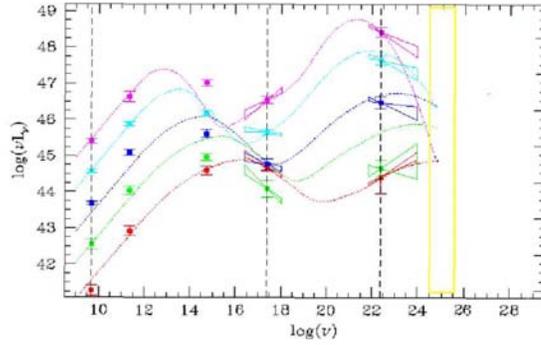


Figure 6: A large sample of blazars (from FSRQ to XBL) has been subdivided within 5 subsamples according to their absolute brightness at radio wave length. Each group is well represented by a single parameter law built upon the hypothesis of associated synchrotron & inverse Compton emissions (see text).

blazars tend to be less bright. Mrk-421 and Mrk-501 data fit well within the overall scheme as part of the utmost energetic group of blazars. The question is now whether the ACT observations can contribute to the comprehension the overall story.

The SSC guide line leaves open the question of the primary electrons. To the least, it permits to deduce from the data, both the energy spectrum of the primary electrons and the magnetic field intensity. The driving force in the case of the Crab plerion was the rotation energy spinning an intense dipole magnet. In the case of AGNs the driving force is the gravitation fall off, and possibly the black hole rotation. There exists models - but not a unique and unambiguous scheme - to generate flares of energetic electrons or protons out of such a machinery. Very likely, it will require many more measurements and tentative modelisations before this question can be mastered. The spectra of the utmost energetic γ -rays is likely to contribute in some essential ways. This is again a specific capability of ACTs to afford precise spectra, particularly during intense flares. A large observational effort is invested into this direction, e.g. the curvature studies of Mrk-421 and Mrk-501 spectra by all the major observatories. In the recent period, at least one new blazar, 1ES1426²³), has been seen by the Whipple and by HEGRA and marginally by CAT, which might also bring a new light on the problem of VHE absorption by the diffuse infrared light. This is the question which we now investigate.

8 Diffuse infrared interstellar absorption

Both Mrk-421 and Mrk-501 are nearby objects, with redshifts of about $z \approx 0.03$. The choice of these targets in the Whipple observational program had been based on this characteristic, not only because of the geometrical d^{-2} flux dependence, but T. Weekes and others were aware of the Universe opacity to VHE γ -rays because of their interaction with diffuse light, with $\gamma\gamma$ conversions into e^+e^- pairs. Necessarily, γ -rays of ≈ 1000 TeV (10^{15} eV) will interact against the cosmological microwave background. This is similar to the GZK effect for protons (Greisen, Zeldovitch and Kuzmin) and constitutes an high energy end point of γ Astronomy. It occurs at lower energies for γ than for protons ($\approx 10^{19}$ eV) because the e^+e^- mass is smaller (by a factor ≈ 100) than the first hadron excitation levels.

Besides the $2.7^\circ K$, the infrared diffuse light from stars and interstellar dusts is likely to affect VHE γ -rays. The deep Universe is transparent in the HE domain, while the ACTs have a limited optical depth. However, for Mrk-421 and Mrk-501, there is today no indication evidences that they are affected by the infrared absorption, no sign of an abrupt cut-off, if we except the above quoted paper of Krennrich et al. on Mrk-501.

The direct measurement of the infrared diffuse light is not easy because of the intense foreground from our own Galaxy (the zodiacal light). The present knowledge combining direct measurements and an evaluation deduced from galaxy counts and star formation hypothesis leads to maximal contributions at both ends, the near IR neighbouring the optical light (directly from stars) and the far IR $2.7^\circ K$ (from star light reprocessed through dust heating). The valley in between is the most problematic region, and this is precisely the region of concern for 10 TeV γ -rays. As no cut-off effect is evidenced, the TeV data can only afford relevant upper limits shown in Fig. 7-left, from Renault et al. ²⁴).

The present situation is very encouraging as it points toward major progress at relatively short terms. The energy region of utmost interest in this context is precisely the unexplored region, in the domain of a few tens of GeV. The blazar being transient sources, the determination of their spectral shape is more difficult as it implies a good sensitivity for a limited time of observation. The number of blazar candidates should increase thanks to a better sensitivity, and to access to lower energy blazars and to larger z -values. This last point is illustrated in Fig 7 showing the optical depth with energy as evaluated by Blanch et

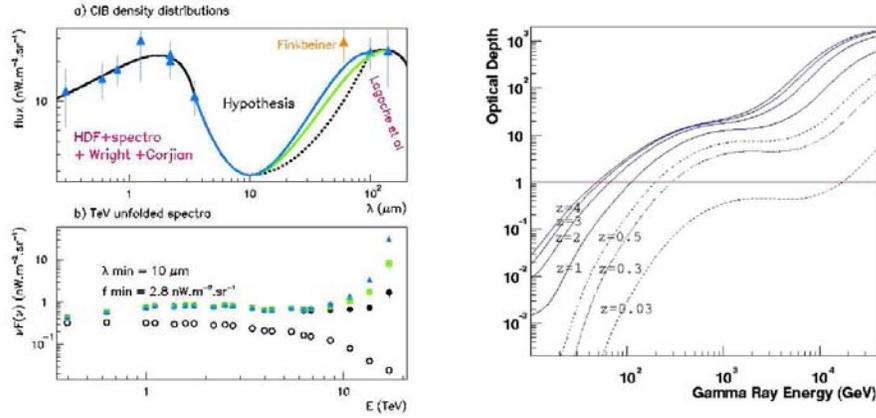


Figure 7: *Left* : The upper figure illustrates the present level of knowledge on the diffuse intergalactic light (in $\text{mW}/\text{mm}^2\text{sr}$), at close infrared (stars) and at far infrared (dusts), with a poorly known valley in between. In the lower figure, the energy spectrum of Mrk 501 in the TeV domain appears as a smoothly decreasing curve in νf_ν . The reconstructed Mrk501 spectra at the source, based on the 3 hypotheses sketched in the upper figure, present awkward rises beyond 10 TeV. This sets an IR upper limit at the level of the dotted line. Similar results were obtained by varying the depth of the valley at $10 \mu\text{m}$. *Right* : based on realistic evaluations of star formation history, the optical depth versus γ energies for different z -values is evaluated. The respective contributions from close and far IR are easily distinguished.

al. 25). The expected cut-off values would be facilitated if the genuine spectral shape was known a priori. This could be settled if an SSC-EC model could be asserted. The information deduced from infrared absorption would then resort not only to the diffuse light in the present Universe but to its evolution down to the z -value of the observed blazars. The absorption will then have the z -dependency of both the VHE γ -rays and diffuse light. If chance permits, these studies could possibly be extended to redshifts $z \geq 2$ thus informing us on the history of star formation which itself should depend on the dark matter constituents.

The very recent evidence for the emission from 1ES1426 is quite important. Its spectrum needs to be studied with great care. Its redshift is $z=0.12$, that is four times as large as that of Mrk-421 and Mrk-501. Much should be understood after fuller analysed and hopefully after more data is collected

(which depends on its level of activity).

9 Conclusions

The ground based γ Astronomy has been for quite a long time just a dream in the mind of adventurous scientists ready to devote their life time for the sake of a new chapter of science, against the disbelief of most of their colleagues. The ACT is now a firmly established means of access to VHE (TeV) γ rays. The evidence for VHE (TeV) γ cosmic sources of various categories is also firmly established. Although the quest for VHE energies started long before the systematic investigations - from space - of HE (GeV) γ sources, it so happened that the two domains of HE & VHE γ Astrophysics reached their golden age in the last decade. However, there remains, in between, an unexplored region ranging from a few GeV to a few hundreds of GeV. It is as if a generation of accelerators (or two) was missing to high energy physicists. With such a remaining gap, no surprise that, from source to source, there is as much contrast as there is an unescapable continuity, as we have seen for instance for the 'school case' of the Crab nebula.

Higher sensitivity instruments will be deployed both in space and on the ground in the next decade. The unexplored region, today the focus of so many interrogations, will then become the best explored region, cumulating information from both classes of observatories, associating wide angular surveys (from space) with well resolved time variations (from ground).

Partly from this totally new window and partly from improved sensitivities at other energies, progress should be at hand for the long lasting question of the origin of CRs and for the modelisation of pulsars, plerions and blazars. Most likely, contributions to other domains will be open, e.g. for the search of dark matter and/or the study of γ ray bursts, not excluding serendipitous discoveries.

alluded to.

References

1. Weekes T.C., et.al. ApJ **342** (1989) 379-395
2. Hoffman C.M. et.al., Rev. Modern Phys., **71** (1999) 897-936. 3(1999) 456

3. Ong, R., Phys.Rep. **305** (1998) 93
4. Gallbraith W. & J.V.Jelley, Nature **171** 349
5. Bacci, C. et al. A&AS..**138** (1999) 597B
6. Amenomori M. et al. ApJ **532** (2000) 302
7. Punch M. et.al. Nature **358** (1992) 477
8. Lebohec S. et al. NIM **416** (1998) 425
9. Tumer, O.T. Nucl.Phys.B (Proc.Suppl.) **14a** (1990) 351
10. Grenier I.A. Nature **404** (2000) 344. & Gherels et al., idem 363
11. Fierro J.M.et.al. in ApJ. **494** (1998) 734-746).
12. Vacanti, G. et al. ApJ **377** (1991) 467
13. Gould, R. PRL **15** (1965) 577
14. de Jager, O.C. & A.K.Harding ApJ **396** (1992) 161
15. Hillas, A.M. et al. ApJ **503** (1998) 744-759
16. de Jager, O.C., Proc. Heidelberg (2000) 154-167. & 613-616
17. Strong, A.W. & I.V.Moskalenko, Ap.J. **509** (1998) 212
18. Aharonian F. et al. A&A **370** (2001) 17-24. & Atoyan A. et al. A&A **354** (2000) 915A
19. Djannati-Attai A. et al. A&A **350** (1999) 17-24
20. Protheroe, R.J. et al., Proc of the 25th ICRC (Durban, S.A.) **8** 317
21. Piron F. et al. A&A in print (and Astroph/0106196)
22. Donato, D., G.Ghisellini, G.Tragliaferri & G.Fossati, ApJ in print (and Astroph/0105203).
23. Communications at the 27th ICRC (Hamburg 2001)
24. Renault, C. et al. A&A in print; (and Astroph/0012094)
25. Blanch, O. & M.Martinez, Astroph/0107582

Science and Technology: The AGILE Scientific Instrument

Guido Barbiellini

Univ. di Trieste and INFN, sezione di Trieste, Italy

Francesco Longo

Univ. di Ferrara and INFN, sezione di Ferrara, Italy

on the behalf of the AGILE collaboration

ABSTRACT

The Energetic Gamma Ray Experiment Telescope was the highest energy instrument on board the Compton Gamma Ray Observatory, and covered the broadest energy range, from 20 MeV to 30 GeV. It had a large field of view, good angular resolution and very low background. Because it was designed for high-energy studies, the detector was optimized to detect gamma rays when they interact by the dominant high-energy pair-production process which forms an electron and a positron pair within the EGRET spark chamber.

The AGILE scientific instrument is based on an innovative design based on three detecting systems: (1) a Silicon Tracker, (2) a Mini-Calorimeter, and (3) an ultralight coded mask system with Si-detectors (Super-AGILE). AGILE is designed to provide: (1) excellent imaging in the energy bands 30 MeV–50 GeV (5–10 arcmin for intense sources) and 10-40 keV (1–3 arcmin);, (2) optimal timing capabilities, with independent readout systems and minimal deadtimes for the Silicon tracker, Super-AGILE and Mini-Calorimeter; (3) large

fields of view for the gamma-ray imaging detector (~ 3 sr) and Super-AGILE (~ 1 sr).

Despite of its smaller dimensions AGILE will have comparable performances to EGRET on axis and substantially better off axis. The innovative technology will allow AGILE to achieve the smallest downtime in high-energy astrophysics.

1 Introduction

The AGILE Mission is the first of the Italian Space Agency Small Scientific Missions ⁸⁾. It is devoted to high-energy astrophysics and is currently planned to be operational in 2003. The AGILE scientific instrument ^{2, 9, 10)} is based on the state-of-the-art technology of solid state Silicon detectors developed by INFN and CNR laboratories. The instrument is light (~ 80 kg) and very effective in detecting and monitoring hard X-ray/gamma-ray sources within a large field of view (FOV).

We adopted the philosophy of one integrated instrument made of three detectors with broad-band detection and imaging capabilities: the Gamma-Ray Imaging Detector (GRID) ⁴⁾ sensitive in the energy range 30 MeV–50 GeV, the hard X-ray imager named Super-AGILE (SA) ⁶⁾ sensitive in the energy range 10–40 keV, and a non-imaging CsI(Tl) Mini-Calorimeter (MC) ¹⁾ sensitive in the energy range 0.3–200 MeV. We briefly describe in these paper the main instrument's characteristics.

The use of the state of the art technology allows AGILE to reach scientific performances better than EGRET.

2 The EGRET experiment

EGRET was sensitive to gamma rays in the energy range from about 30 MeV to 30 GeV. In the mode used for most of the observations, the effective area of the telescope is about 1000 cm² at 150 MeV, 1500 cm² around 0.5-1 GeV, decreasing gradually at high energies to about 700 cm² at 10 GeV for targets near the center of the field of view. EGRET's effective area is maximum when the target is on axis and falls to approximately 50% of this value when the angular offset reaches 18°.

The instrument had components typically used in the high energy gamma-ray telescopes until the 1990's; an anticoincidence system to discriminate against charged particle radiation, a multilevel thin-plate spark chamber system to convert gamma rays and determine the trajectories of the secondary electron-positron pair, a triggering telescope that detects the presence of the pair with

the correct direction of the motion, and an energy measuring calorimeter, which in the case of EGRET is a NaI(Tl) crystal. Descriptions of the instrument and details of the instrument calibration, both before and after launch, could be found in [12, 7]. The instrument was carefully designed to be essentially free of internal background, and calibration tests verified that the internal background was at least an order of magnitude below the extragalactic diffuse gamma radiation.

The scientific goals of the mission included the study of the high energy transfers in neutron stars, other galactic objects, and active galaxies, the galactic and extragalactic high-energy gamma-ray diffuse radiation, energetic solar phenomena, cosmic rays and supernovae, and the high-energy gamma-ray emission of the gamma ray bursts.

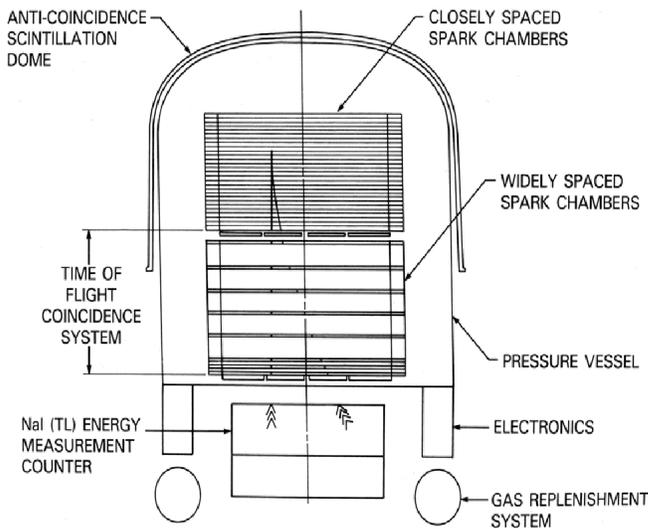


Figure 1: *Schematic view of the EGRET instrument*

3 The AGILE Instrument

Fig. 2 shows the AGILE instrument configuration of total weight of ~ 80 kg including the Si-Tracker, Super-AGILE, Mini-Calorimeter, the Anticoincidence system and electronics. The baseline AGILE instrument is made of the following elements.

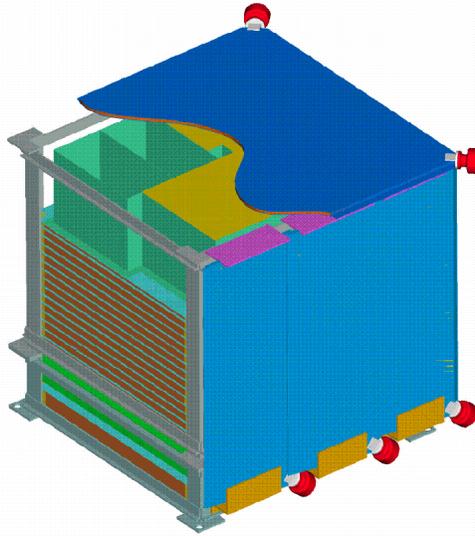


Figure 2: *Schematic view of the AGILE instrument*

- **Silicon-Tracker**, a gamma-ray pair-converter and imager made of 14 planes, with two Si-layers per plane providing the X and Y coordinates of interacting charged particles. The fundamental Silicon detector unit is a tile of area $9.5 \times 9.5 \text{ cm}^2$, microstrip pitch equal to $121 \mu\text{m}$, and thickness $410 \mu\text{m}$. The adopted “floating readout strip” system has a total of 384 readout channels (readout pitch equal to $242 \mu\text{m}$) and three readout TA1 chips per Si-tile. Each Si-Tracker layer is made of 4×4 tiles, for a total geometric area of $38 \times 38 \text{ cm}^2$ and 1,536 readout channels. The first 12 planes are made of three elements: a first layer of Tungsten ($0.07 X_0$) for gamma-ray conversion, and two Si-layers (views) with microstrips orthogonally positioned. For each plane there are then $2 \times 1,536$ readout microstrips. Since the GRID trigger requires at least three Si-planes to be activated, two more Si-planes are inserted at the bottom of the Tracker without Tungsten layers. The total readout channel number of for the GRID Tracker is $\sim 43,000$. Both digital and analog information (charge deposition in Si-microstrip) is read by TA1 chips. These channels are individually read by state-of-the-art electronic devices (front-end-electronics, FEE). The distance between mid-planes equals 1.6 cm (optimized by Montecarlo simulations). Special algorithms applied off-line to telemetered data will allow optimal background subtraction and reconstruction of the photon incidence angle. Both digital and analog information are crucial for this task. The positional resolution obtained by these detec-

tors in recent beam tests at CERN is excellent, being below $40 \mu\text{m}$ for a large range of photon incidence angles ³⁾. More information on the Silicon Tracker can be found in Refs. ⁴⁾.

- **Super-AGILE**, made of four square Silicon detectors ($19 \times 19 \text{ cm}^2$ each) and associated FEE placed on the first GRID tray plus an ultra-light coded mask system supporting a Tungsten mask placed at a distance of 14 cm from the Silicon detectors. Super-AGILE tasks are: *(i)* photon-by-photon detection and imaging of sources in the energy range 10-40 keV, with a field-of-view (FOV) of $\gtrsim 0.8 \text{ sr}$, good angular resolution (1-3 arcmins, depending on source intensity and geometry), and good sensitivity ($\sim 5 \text{ mCrab}$ for 50 ksec integration, and $\lesssim 1 \text{ Crab}$ for a few seconds integration); *(ii)* simultaneous X-ray and gamma-ray spectral studies of high-energy sources; *(iii)* excellent timing ($\lesssim 4 \mu\text{s}$); *(iv)* burst trigger for the GRID and MC; *(v)* GRB alert and quick on-board positioning capability. Refs. ⁶⁾ describe the Super-AGILE structure and scientific capabilities.
- **Mini-Calorimeter (MC)**, made of two planes of Cesium Iodide (CsI) bars, for a total (on-axis) radiation length of $1.5 X_0$. The signal from each CsI bar is collected by two photodiodes placed at both ends. The MC tasks are: *(i)* obtaining additional information on the energy of particles produced in the Si-Tracker; *(ii)* detecting GRBs and other impulsive events with spectral and intensity information in the energy band $\sim 0.3 - 100 \text{ MeV}$. We note that the problem of “particle backsplash” for AGILE is much less severe than in the case of EGRET. AGILE allows a relatively efficient detection of (inclined) photons near 10 GeV and above also because the AC-veto can be disabled for events with more than $\sim 100 \text{ MeV}$ total energy collected in the MC. Ref. ¹⁾ describes the MC characteristics.
- **Anticoincidence System**, aimed at both charged particle background rejection and preliminary direction reconstruction for triggered photon events. The AC system surrounds all AGILE detectors (Super-AGILE, Si-Tracker and MC). Each lateral face is segmented with three plastic scintillator layers (0.6 cm thick) connected to photomultipliers placed at their bottom. A single square plastic scintillator layer (0.5 cm thick) constitutes the top-AC layer whose signal is read by four photomultipliers placed at the four corners.
- **Data Handling System**, for fast processing of the GRID, Mini-Calorimeter and Super-AGILE events. The GRID trigger logic for the acquisition of gamma-ray photon data and background rejection is structured in two

main levels: Level-1 and Level-2 trigger stages. The Level-1 trigger is fast ($\lesssim 5\mu\text{s}$) and requires a signal in at least three out of four contiguous tracker planes, and a proper combination of fired TA1 chip number signals and AC signals. An intermediate Level-1.5 stage is also envisioned (lasting $\sim 20\mu\text{s}$), with the acquisition of the event topology based on the identification of fired TA1 chips. Both Level-1 and Level-1.5 have a hardware-oriented veto logic providing a first cut of background events. Level-2 data processing includes a GRID readout and pre-processing, “cluster data acquisition” (analog and digital information), and processing by a dedicated CPU. The Level-2 processing is asynchronous (estimated duration \sim a few ms) with the actual GRID event processing. The GRID deadtime turns out to be $\sim 100\mu\text{s}$ and is dominated by the Tracker readout.

The charged particle and albedo-photon background passing the Level-1+1.5 trigger level of processing is simulated to be $\lesssim 100$ events/sec for the nominal equatorial orbit of AGILE⁵). The on-board Level-2 processing has the task of reducing this background by a factor between 3 and 5. Off-line processing of the GRID data with both digital and analog information is being developed with the goal to reduce the particle and albedo-photon background rate above 100 MeV to ~ 0.01 events/sec.

In order to maximize the GRID FOV and detection efficiency for large-angle incident gamma-rays (and minimize the effects of particle back-splash from the MC and of “Earth albedo” background photons), the data acquisition logic uses proper combinations of top and lateral AC signals and a coarse on-line direction reconstruction in the Si-Tracker. For events depositing more than ~ 100 MeV in the MC, the AC veto can be disabled to allow the acquisition of gamma-ray photon events with energies larger than 1 GeV.

Appropriate data buffers and burst search algorithms are envisioned to maximize data acquisition for transient gamma-ray events (e.g., GRBs) in the Si-Tracker, Super-AGILE and Mini-Calorimeter, respectively.

The Super-AGILE event acquisition is conceptually simple. After a first “filtering” based on AC-veto signals and pulse-height discrimination in the dedicated FEE (XAA1 chips), the events are buffered and transmitted to the CPU for burst searching and final data formatting. The 4 Si-detectors of Super-AGILE are organized in 16 independent readout units, of $\sim 5\mu\text{s}$ deadtime each.

Given the relatively large number of readable channels in the Si-Tracker and Super-AGILE ($\sim 50,000$ channels), the instrument requires a very efficient readout system. In order to maximize the detecting area and minimize the instrument weight and absorbed power, the GRID and Super-

AGILE front-end-electronics is partly accommodated in special boards placed externally on the Tracker lateral faces. Electronic boxes, P/L memory (and buffer) units will be accommodated at the bottom of the instrument. Ref. ¹¹⁾ describes the AGILE Data Handling System.

Table 1 summarizes the main characteristics of the AGILE gamma-ray instrument and its performance compared to that of EGRET. We assumed a typical 2-week pointing duration and a $\sim 50\%$ exposure efficiency.

Table 1: **A COMPARISON BETWEEN EGRET AND AGILE**

	EGRET	AGILE
Mass	1830 kg	80 kg
Gamma-ray energy band	30 MeV – 30 GeV	30 MeV–50 GeV
Field of View	~ 0.5 sr	~ 3 sr
PSF	5.5°	4.7° (@ 0.1 GeV)
(68% containment radius)	1.3°	0.6° (@ 1 GeV)
	0.5°	0.2° (@ 10 GeV)
Deadtime for γ -ray detection	$\gtrsim 100$ ms	$\lesssim 100$ μ s
Sensitivity	8×10^{-9}	6×10^{-9} (@ 0.1 GeV)
for pointlike sources	1×10^{-10}	4×10^{-11} (@ 1 GeV)
(ph cm $^{-2}$ s $^{-1}$ MeV $^{-1}$)	1×10^{-11}	3×10^{-12} (@ 10 GeV)
Required pointing reconstruction	~ 10 arcmin	~ 1 arcmin

4 Conclusions

The AGILE scientific instrument is innovative in many ways, and is designed to obtain an optimal gamma-ray detection performance despite its relatively small mass and absorbed power. The refined readout of the Silicon Tracker allows to reach an excellent spatial resolution ($\sim 40 \mu\text{m}$) that is crucial for gamma-ray imaging. The combination of hard X-ray (Super-AGILE) and gamma-ray imaging capabilities in a single integrated instrument is unique to AGILE. We anticipate a crucial role of Super-AGILE for studies of AGNs, GRBs, and Galactic sources. Positioning better than ~ 6 arcmin can be obtained for sources detectable in the hard X-ray range. Instrumental deadtimes for the different detectors are unprecedentedly small for gamma-ray instruments, and microsecond photon timing can be achieved. An optimal Burst Search Procedure is implemented in the on-board Data Handling System allowing a GRB

search for a broad dynamic range of durations from milliseconds to hundreds of seconds.

References

1. Auricchio N. et al., *Proceedings of Gamma2001 Conference*, AIP Conf. Proceedings, eds. Ritz S., Gehrels N. and Shrader C.R., 2001, Vol. 587, p. 749
2. Barbiellini G. et al., *Proceedings of the 5th Compton Symposium*, AIP Conf. Proceedings, ed. M. McConnell, 2001, Vol. 510, p. 750
3. Barbiellini G. et al., *NIM-A*, 2001, submitted
4. Barbiellini G. et al., *Proceedings of Gamma2001 Conference*, AIP Conf. Proceedings, eds. Ritz S., Gehrels N. and Shrader C.R., 2001, Vol. 587, p. 754
5. Cocco V., Longo F. & Tavani M., *Proceedings of Gamma2001 Conference*, AIP Conf. Proceedings, eds. Ritz S., Gehrels N. and Shrader C.R., 2001, Vol. 587, p. 744
6. Lapshov I. et al., *Proceedings of Gamma2001 Conference*, AIP Conf. Proceedings, eds. Ritz S., Gehrels N. and Shrader C.R., 2001, Vol. 587, p. 769
7. Nolan P.L. et al., 1992, *IEEE Transactions Nucl. Sci.*, 39, 993
8. Tavani M. et al., *The AGILE Mission*, 2001 <http://www.ifctr.mi.cnr.it/Agile>
9. Tavani M. et al., *Proceedings of the 5th Compton Symposium*, AIP Conf. Proceedings, ed. M. McConnell, 2001, Vol. 510, p. 746
10. Tavani M. et al., *Proceedings of Gamma2001 Conference*, AIP Conf. Proceedings, eds. Ritz S., Gehrels N. and Shrader C.R., 2001, Vol. 587, p. 729
11. Tavani M. et al., *NIM-A*, 2001, in preparation
12. Thompson D.J. et al., 1993, *ApJS*, 86, 629

GAMMA-RAY ASTROPHYSICS AND THE AGILE MISSION

Marco Tavani

Istituto Fisica Cosmica, CNR, via Bassini 15, I-20233 Milano (Italy)

ABSTRACT

Gamma-rays of cosmic origin above nuclear transition energies (larger than 10 MeV) are a manifestation of remarkable energetic acceleration and radiative processes. Diffuse gamma-ray radiation from our Galaxy and about three hundred sources reveal a very active Universe in the energy range 30 MeV – 20 GeV. Seven pulsars and about sixty-five Active Galactic Nuclei are currently identified with gamma-ray sources, and five gamma-ray bursts were clearly detected above a few tens of MeV. However, many more sources (~ 200) are still unidentified, and their origin and underlying physical processes are a challenge to theoretical models. We briefly review the main astrophysics issues (both theoretical and observational) to be addressed in the following years, and how the planned AGILE mission will contribute to resolve many open issues.

1 Introduction

Observational gamma-ray astronomy is a relatively young branch of astrophysics. The first certain satellite detection of photons above 50 MeV (from the Galactic disk) was made by the third Orbiting Solar Observatory (OSO-3) in the late Sixties²⁵⁾, together with high-altitude balloon experiments' detections of the Crab pulsar⁸⁾, and diffuse gamma-ray radiation from the Galaxy²³⁾. Improved instrumentation on board of the SAS-2 satellite (Nov. 1972 – July 1973) provided a first sky map of the Galaxy and detection of several sources¹⁵⁾. The European Cosmic-Ray Satellite (COS-B, August 1975 – April 1982) concentrated in observing the Galactic plane with only a few pointings at high Galactic latitudes. About 20 gamma-ray sources could be clearly detected near the plane, together with the first extragalactic source (3C273)^{7, 32)}. A great advance came with the results of the EGRET instrument^{16, 37)} on board of the Gamma-Ray Observatory (GRO) (April 1991 – June 2000). About 300 gamma-ray sources above 30 MeV were detected including about 60 extragalactic ones^{20, 38, 39, 40)}. Currently, seven isolated pulsars are identified as sources of pulsed gamma-ray emission⁴⁰⁾, and the big surprise of EGRET, without any doubt, came from the discovery of strongly variable gamma-ray emission from blazars (a special class of Active Galactic Nuclei¹⁹⁾). In addition, important results were obtained on the cosmic-ray origin³⁰⁾, diffuse Galactic²¹⁾ and extragalactic³¹⁾ emissions, and gamma-ray burst detections^{12, 22)}.

The inheritance of past space missions capable of detecting cosmic radiation above 30 MeV is then remarkable. The sky shines in gamma-rays in ways that could not be anticipated before observations and challenging theoretical models. In the following, we briefly outline the main open questions with a perspective about the future.

2 Open Issues

Gamma-rays above 30 MeV provide a crucial diagnostic for a variety of fundamental topics dealing with the most energetic processes in our Universe, including the origin of cosmic-rays, the emission from magnetized neutron stars and black holes, the behavior and jet processes of Active Galactic Nuclei and GRB sources.

2.1 Acceleration processes

A crucial, if not *the* crucial, topic of high-energy astrophysics is particle acceleration. Particle energy distribution functions are usually assumed to be power-laws. However, we need to understand in more detail the fundamental processes of acceleration for a broad variety of environments and boundary conditions. Diffusive Fermi-like acceleration^{13, 14)} by hydromagnetic turbulence or parallel-shocks has important applications in cosmic-ray acceleration in supernova-remnants and possibly AGN jets. Also resonant scattering and acceleration by magneto-hydrodynamical and plasma effects can also have relevant applications especially when parallel-shock conditions do not apply^{41, 42)}. Alternately, strong electric fields in plasmas locally avoiding strong e^+/e^- creation are relevant for pulsar magnetospheres, and possibly jets and relativistic ejecta from compact objects.

Gamma-ray observations challenge theoretical models in a variety of ways: particle composition, acceleration processes and timescales, efficiencies, and the competition with gyro-synchrotron and inverse Compton cooling are under discussion for both Galactic and extragalactic sources. Despite a large number of models, no consensus is reached on a variety of fundamental topics. In order to resolve the *empasse*, we need higher statistics and time resolution for photon detection together with a refinement of theoretical models.

2.2 Cosmic rays and diffuse Galactic gamma-ray emission

Diffuse gamma-ray emission from the Galaxy dominates the detected flux. These diffuse photons are a manifestation of cosmic-ray propagation and bombardment in gaseous environments, and are also produced by synchrotron and inverse Compton processes²¹⁾. Gamma-rays reflect the spiral and local geometry of the Galaxy depending on the number density of cosmic-rays (that may vary across the Galaxy because of source localization) and the gas distribution. Photons above 30 MeV are therefore a crucial diagnostic for cosmic-ray processes.

EGRET contributed in an important way to the issue of the Galactic or extragalactic origin of cosmic rays. The lack of detection from the Small Magellanic Cloud at the level below what expected for an extragalactic origin (upper limit of 0.5×10^{-7} ph. cm⁻² s⁻¹ above 100 MeV) suggests that cosmic-rays originate in our Galaxy²⁹⁾. Acceleration of hadronic cosmic-rays in supernova

ejecta might produce a detectable flux of gamma-rays in nearby supernova remnants (SNRs), but at the moment we are still missing the smoking gun. TeV emission from the Crab ¹¹⁾ and SN1006 ²⁴⁾ indicate the existence of relativistic leptonic populations emitting by inverse Compton scattering. However, there is no proof yet of the association of a SNR and high energy emission originating from accelerated hadrons. Despite several interesting candidates, no Galactic supernova remnant could be unambiguously associated with a unidentified EGRET source because of limited spatial resolution. This issue will definitely be resolved by next generation gamma-ray detectors in space, providing a substantially better localization compared to that of EGRET.

2.3 Pulsar physics

Seven isolated pulsars were detected by EGRET providing a remarkable set of data ⁴⁰⁾. Precise timing of gamma-rays is difficult because of a limited statistics and pulsar (micro) glitches that limit the phase coherence reconstruction. Searches of pulsed gamma-ray signals in the EGRET database is therefore difficult without hints from other wavelengths (e.g., radio, X-rays).

The future is promising. Radio monitoring together with gamma-ray larger exposures on fields containing unidentified gamma-ray sources, improved spatial resolution, and better gamma-ray timing properties should lead to new pulsar discoveries. The current debate is on competing pulsar emission models (polar cap vs. outer gap), the existence of a radio-less population of Geminga-like gamma-ray pulsars, whether millisecond pulsars are detectable in gamma-rays, and time-resolved spectral and light curve features. New radio data from the multi-beam Parkes survey ⁹⁾ led to the discovery of about 30 new young pulsars (age less than 10^5 years). Some of them are apparently coincident with unidentified EGRET sources ¹⁰⁾. Clearly, these young pulsars with characteristics similar to known gamma-ray pulsars are ideal candidates for pulsed gamma-ray detection by future instruments.

2.4 Active Galactic Nuclei: engines and jets

EGRET remarkable discovery of strongly variable gamma-ray emission from blazars started a new way of investigating black holes and jets. Different "states" of core emission and X-ray/gamma-ray/TeV jet emission were detected for a variety of blazars by combining multifrequency observations. Gamma-ray

flaring activity appears to be unpredictable, and apparently lasting days-weeks in our observer's frame. Practically all blazars detected by EGRET are strongly variable. In a few occasions, several high-states above 100 MeV were detected from the same source (e.g., 3C279) for an apparent duty cycle of a few years for large flares. The understanding of particle acceleration and evolution in the jet is still preliminary, and certainly we need more exposure, photon statistics, and a larger sample of detected AGNs to study in detail the fundamental processes. What is the ultimate origin of gamma-ray flares ? Is there any relation with the emission of relativistic radio plasmoids ? What is the blazar gamma-ray duty cycle ? What are the particle acceleration and cooling processes in the jet ? What is the physical difference between blazars emitting in the GeV energy range from those emitting in the TeV range ? Is the extragalactic diffuse gamma-ray background resolvable only in terms of AGNs ? These are only a few questions that can be addressed by future investigations of the AGN population.

2.5 Gamma-ray bursts

Relativistic dynamics and acceleration/cooling processes in jets can have important applications for the interpretation of gamma-ray bursts (GRBs) high-energy emission. EGRET observations of GRB emission above 30 MeV provide a challenge to theoretical models. Impulsive prompt emission of gamma-ray photons within instrumental deadtimes (100 ms), absence of a spectral cut-off up to 10 GeV, durations substantially longer than those at lower energies indicating the clear existence of "delayed (afterglow) gamma-ray emission", spectral variability and re-acceleration in multiple pulse events all demonstrate the complexity of the engine, and the realization of remarkable particle acceleration and radiative efficiencies in GRBs. Only 5 GRBs were detected by the EGRET spark chamber in approximately 5 years of operations ¹²⁾. Clearly, we need more.

3 Needs

OSO-3, SAS-2, COS-B, and EGRET were remarkable instruments, each contributing an important piece of the puzzle that we are trying to compose. However, if we want to make progress, we need a substantial improvement of

the scientific performance of gamma-ray instruments. Based on the accumulated evidence and the technical development of gamma-ray detectors, we are in desperate needs of:

- **better angular resolution and larger fields of view** (an apparently contradictory request !), improving EGRET error boxes by a factor of at least 4 in sky area, and making the FOV larger by a factor of 4-5;
- **better timing**, reaching a few microsecond for photon tagging and a deadtime for gamma-ray detection below 1 ms;
- **capability of simultaneous X-ray and gamma-ray detection by the same instrument**, to improve the physical information on processes involving broad-band emission, and to provide error boxes of order of 2-3 arcmin.

In the following, we will discuss the first of future gamma-ray missions with these capabilities.

4 The AGILE Mission

The space program AGILE (*Astro-rivelatore Gamma a Immagini LEggero*) is planned to be the first of the ASI Scientific Small Missions ^{34, 35}. AGILE will be the only Mission entirely dedicated to gamma-ray astrophysics (30 MeV–50 GeV) during the period 2003-2006. AGILE is currently in Phase C ⁴). The AGILE scientific instrument is based on the state-of-the-art and reliably developed technology of solid state Silicon detectors developed by the Italian INFN laboratories ^{1, 2, 3, 18}). The instrument is relatively light (~ 100 kg) and effective in detecting and monitoring gamma-ray sources within a large field of view.

AGILE's philosophy is to have one integrated instrument made of three detectors with broad-band detection and imaging capabilities. The Gamma-Ray Imaging Detector (GRID) is sensitive in the energy range ~ 30 MeV–50 GeV. It will be characterized by the smallest ever obtained deadtime for gamma-ray detection ($\lesssim 100 \mu\text{s}$) and by a trigger based exclusively on Silicon plane detectors. The GRID consists of a Silicon-Tungsten Tracker, a Cesium Iodide Mini-Calorimeter, an Anticoincidence system made of segmented plastic

scintillators, and fast readout electronics and processing units. The GRID is designed to achieve an optimal angular resolution (source location accuracy $\sim 5' - 20'$ for intense sources), an unprecedentedly large field-of-view (~ 3 sr), and a sensitivity comparable to that of EGRET for on-axis (and substantially better for off-axis) point sources.

AGILE will also have detection and imaging capabilities in the hard X-ray range provided by the Super-AGILE detector. It consists of an additional plane of four Silicon square detectors positioned on top of the GRID Tracker plus an ultra-light coded mask structure whose top absorbing mask is at the distance of 14 cm from the silicon detectors. The main goals of Super-AGILE are the simultaneous gamma-ray and hard X-ray detection of astrophysical sources (unprecedented for gamma-ray instruments), optimal source positioning (1-3 arcmins, depending on intensity), fast burst alert and on-board trigger capability.

The CsI Mini-Calorimeter (MC) will also detect and collect events independently of the GRID. The energy range for this non-imaging detector is 0.3–200 MeV, and it can be very useful to provide spectral and accurate timing information of transient events. The content of a cyclic MC event buffer will be transmitted to the ground for impulsive events (solar flares, GRBs, other transients).

AGILE with its combination of GRID, MC, and Super-AGILE is a very innovative instrument, with an optimal expected performance for transients (GRBs, stellar flares, unidentified gamma-ray sources, AGNs) and steady sources (e.g., pulsars). The fast AGILE electronic readout and data processing (resulting in very small detectors' deadtimes) allow for the first time the systematic search for sub-millisecond gamma-ray transients³⁶⁾ with durations comparable with the dynamical timescale of $\sim 1 M_{\odot}$ compact objects.

It is clear today that successful investigations of gamma-ray sources rely on coordinated space and ground-based observations. The AGILE Science Program will be focused on a prompt response to gamma-ray transients and alert for follow-up multiwavelength observations. AGILE will provide crucial information complementary to the many space missions that will be operational during the first decade of the new Millennium (INTEGRAL, XMM, CHANDRA, SWIFT, and others). Furthermore, it can support ground-based investigations in the radio, optical, and TeV bands. No other mission entirely dedicated to

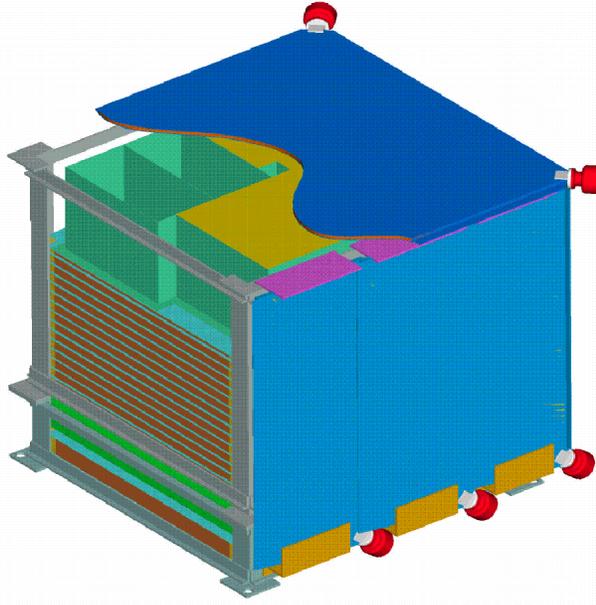


Figure 1: *Schematic view of the AGILE instrument (AC System partially displayed). The GRID is made of a Silicon Tracker (14 Tungsten and Silicon planes) and a Mini-Calorimeter placed at the bottom of the instrument. Super-AGILE has its 4 Si-detectors placed at the top of the first GRID tray, and an ultra-light coded mask system (CMS) positioned on top (the figure shows the CMS partition configuration). The instrument size is $\sim 63 \times 63 \times 58.5 \text{ cm}^3$, including Super-AGILE and the AC System for a total weight of $\sim 100 \text{ kg}$.*

gamma-ray astrophysics above 30 MeV is being planned before GLAST. The technological and scientific development of AGILE is also strongly integrated towards GLAST. Part of the AGILE Science Program will be open for Guest Investigations on a competitive basis. Quicklook data analysis and fast communication of new transients will be implemented as an essential part of the AGILE Science Program.

5 Science with AGILE

5.1 Gamma-Ray Astrophysics with the GRID

The GRID has been designed to obtain:

- **excellent imaging capability in the energy range 100 MeV-50 GeV**, improving the EGRET angular resolution by a factor of 2 (see Fig. 2);
- **a very large field-of-view**, allowing simultaneous coverage of $\sim 1/4$ of the entire sky per each pointing (FOV larger by a factor of ~ 6 than that of EGRET);
- **excellent timing capability**, with absolute time tagging of uncertainty near $1 \mu\text{s}$ and very small deadtimes ($\sim 100 \mu\text{s}$ for the Si-Tracker and $\sim 20 \mu\text{s}$ for each of the individual CsI bars);
- **a good sensitivity for point sources**, comparable to that of EGRET for *on-axis* sources, and substantially better for *off-axis* sources;
- **excellent sensitivity to photons in the energy range $\sim 30\text{-}100$ MeV**, with an effective area above 200 cm^2 at 30 MeV;
- **a very rapid response to gamma-ray transients and gamma-ray bursts**, obtained by a special quicklook analysis program and coordinated ground-based and space observations.

5.1.1 Large FOV monitoring of gamma-ray sources

Fig. 3 show a typical AGILE pointing. Relatively bright AGNs and Galactic sources flaring with fluxes larger than $10^{-6} \text{ ph cm}^{-2} \text{ s}^{-1}$ (above 100 MeV) can be detected within a few days by the AGILE quicklook analysis. We conservatively estimate that for a 3-year mission AGILE is potentially able to detect a number of gamma-ray flaring AGNs larger by a factor of several compared to that obtained by EGRET during its 6-year mission. Furthermore, the large FOV will favor the detection of fast transients such as gamma-ray bursts. Taking into account the high-energy distribution of GRB emission above 30 MeV, we conservatively estimate that ~ 1 GRB/month can be detected and imaged in the gamma-ray range by the GRID.

5.1.2 Fast reaction to strong high-energy transients

The existence of a large number of variable gamma-ray sources (extragalactic and near the Galactic plane ³³) makes necessary a reliable program for quick

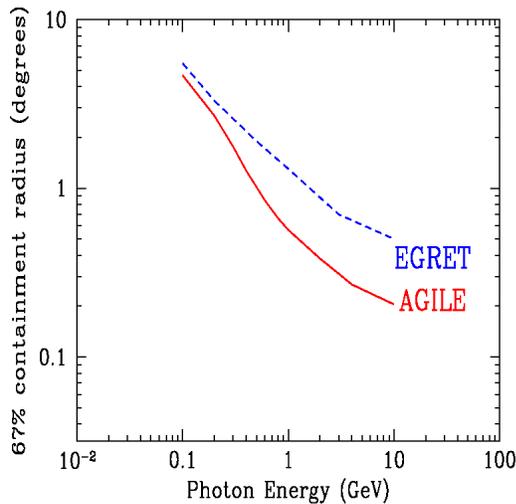


Figure 2: *Three dimensional PSF (67% containment radius) as a function of photon energy for AGILE-GRID and EGRET.*

response to transient gamma-ray emission. Quicklook analysis of gamma-ray data is a crucial task to be carried out by the AGILE Team. Prompt communication of gamma-ray transients (that require typically 2-3 days to be detected with high confidence for sources above 10^{-6} ph cm $^{-2}$ s $^{-1}$) will be ensured. Detection of short timescale (seconds/minutes/hours) transients (GRBs, SGRs, solar flares and other bursting events) is possible in the gamma-ray range. A primary responsibility of the AGILE Team will be to provide positioning of short-timescale transient as accurate as possible, and to alert the community through dedicated channels.

5.1.3 Large exposures for Galactic and extragalactic sky regions

The AGILE average exposure per source will be larger by a factor of ~ 4 for a 1-year sky-survey program compared to the typical exposure obtainable by EGRET for the same time period. After a 1-year all-sky pointing program, AGILE average sensitivity to a generic gamma-ray source above the Galactic plane is expected to be better than EGRET by a factor conservatively given as

tation of the electronic readout of MC and Super-AGILE detectors (32 MC elements and 16 Super-AGILE elements) the effective deadtimes will be much less than those for individual units.

Furthermore, the MC events detected during the Si-Tracker readout deadtime will be automatically stored in the GRID event. For these events, precise timing and detection in the $\sim 1\text{--}200$ MeV range can be achieved with temporal resolution well below $100\ \mu\text{s}$. This is crucial for AGILE high-precision timing investigations.

Fig. 4 show the AGILE deadtime performance compared to other gamma-ray missions. Fast AGILE timing will, for the first time, allow investigations and searches for sub-millisecond transients in the gamma-ray energy range.

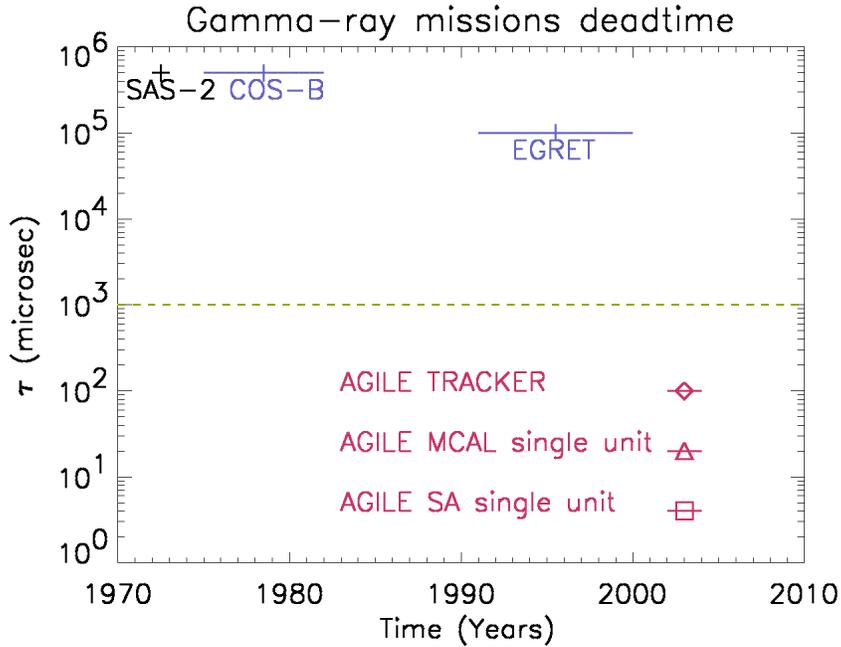


Figure 4: Instrumental deadtimes (τ) for the AGILE detectors and previous gamma-ray instruments.

5.2 Super-AGILE

An imaging coded mask detector system (Super-AGILE) in addition to the GRID will provide a unique tool for the study of high-energy sources. The Super-AGILE FOV is planned to be ~ 0.8 sr. Super-AGILE can provide important information including:

- **source detection and spectral information in the energy range $\sim 10\text{-}40$ keV** to be obtained simultaneously with gamma-ray data (5 mCrab sensitivity at 15 keV (5σ) for a 50 msec integration time);
- **accurate localization ($\sim 1\text{-}2$ arcmins) of GRBs and other transient events** (for typical transient fluxes above ~ 1 Crab); the expected GRB detection rate is $\sim 1 - 2$ per month;
- **excellent timing**, with absolute time tagging uncertainty and deadtime near $4\ \mu\text{s}$ for each of the 16 independent readout units of the Super-AGILE Si-detector;
- **long-timescale monitoring (~ 2 weeks) of hard X-ray sources;**
- **hard X-ray response to gamma-ray transients detected by the GRID**, obtainable by slight repointings of the AGILE spacecraft (if necessary) to include the gamma-ray flaring source in the Super-AGILE FOV.

The combination of simultaneous hard X-ray and gamma-ray data will provide a formidable combination for the study of high-energy sources. Given the sensitivities of the GRID and Super-AGILE, simultaneous hard X-ray/gamma-ray information is anticipated to be obtainable for: (1) GRBs, (2) blazars with strong X-ray continuum such as 3C 273 and Mk 501, (3) Galactic jet-sources with favorable geometries, (4) unidentified variable gamma-ray sources. Figs. 5,6,7 show the expected scientific performance of Super-AGILE.

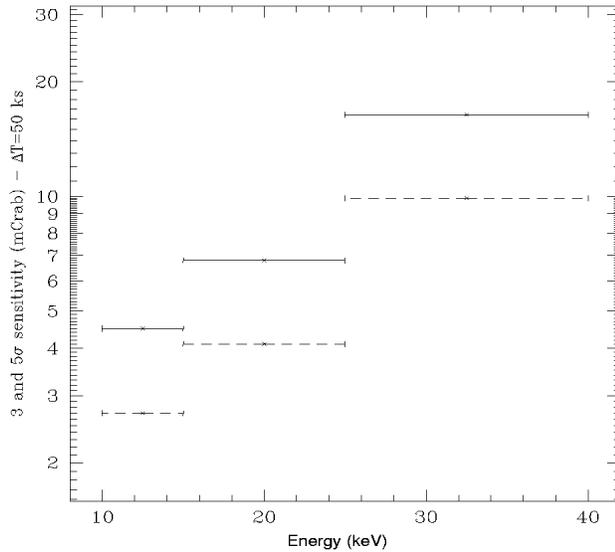


Figure 5: *Super-AGILE* simulated sensitivity (solid data points: 5σ , dashed data points: 3σ) for a 50 ksec integration in mCrab units.

5.3 Scientific Objectives

We summarize here the main AGILE's scientific objectives (listed without any meaning to the ordering, Table 3 provides a schematic summary).

- **Active Galactic Nuclei.** For the first time, simultaneous monitoring of a large number of AGNs per pointing will be possible. Several outstanding issues concerning the mechanism of AGN gamma-ray production and activity can be addressed by AGILE including: (1) the study of transient vs. low-level gamma-ray emission and duty-cycles; (2) the relationship between the gamma-ray variability and the radio-optical-X-ray-TeV emission; (3) the correlation between relativistic radio plasmoid ejections and gamma-ray flares; (4) hard X-ray/gamma-ray correlations. A program for joint AGILE and ground-based

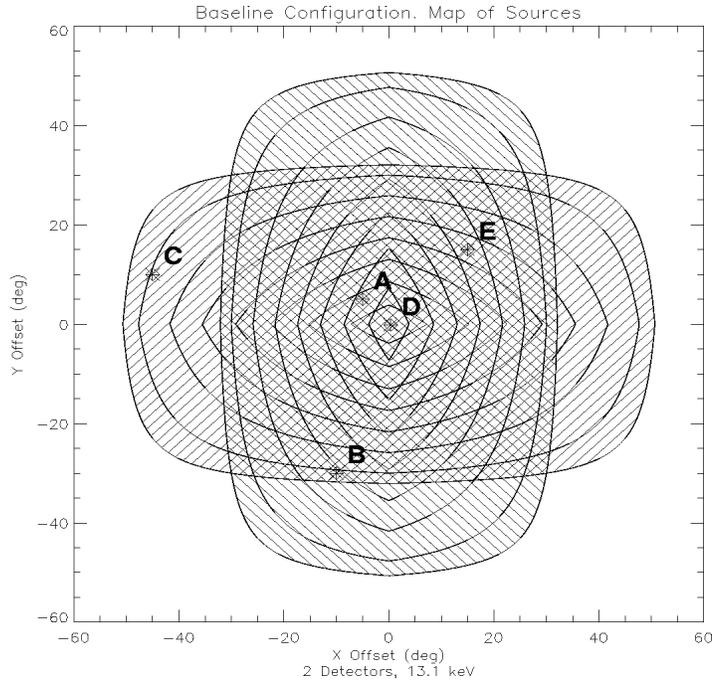


Figure 6: *The Super-AGILE field of view, showing the overlap between sky regions covered by the orthogonal one-dimensional coded mask units.*

monitoring observations is being planned. On the average, AGILE will achieve deep exposures of AGNs and substantially improve our knowledge on the low-level emission as well as detecting flares. We conservatively estimate that for a 3-year program AGILE will detect a number of AGNs 2–3 times larger than that of EGRET. Super-AGILE will monitor, for the first time, simultaneous AGN emission in the gamma-ray and hard X-ray ranges.

- **Gamma-ray bursts.** About ten GRBs were detected by the EGRET spark chamber during ~ 7 years of operations ²⁷). This number was limited by the EGRET FOV and sensitivity and not by the GRB emission mechanism. GRB detection rate by the GRID is expected to be at least a factor of ~ 5 larger than that of EGRET, i.e., ≥ 5 –10 events/year). The small GRID deadtime (~ 1000 times smaller than that of EGRET) allows a better study of the initial phase of GRB pulses (for which EGRET response was in many cases

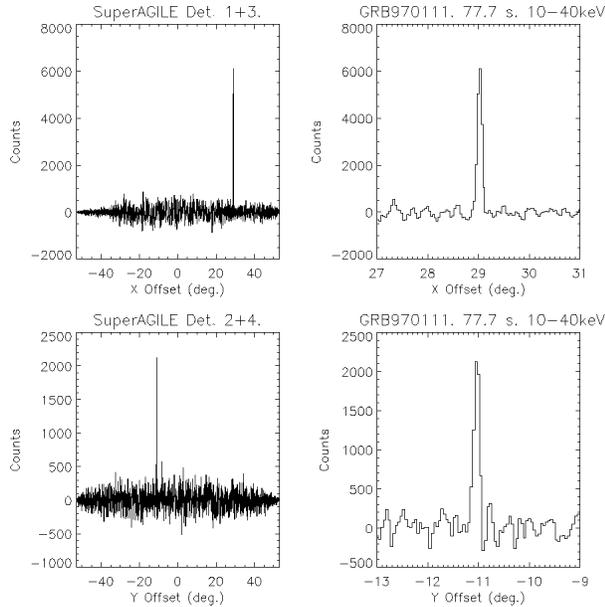


Figure 7: *Simulated detection of GRB 970111 by Super-Agile.*

inadequate). The remarkable discovery of ‘delayed’ gamma-ray emission up to ~ 20 GeV from GRB 940217²²⁾ is of great importance to model burst acceleration processes. AGILE is expected to be highly efficient in detecting photons above 10 GeV because of limited backscattering. Super-AGILE will be able to locate GRBs within a few arcminutes, and will systematically study the interplay between hard X-ray and gamma-ray emissions. Special emphasis is given to fast timing allowing the detection of sub-millisecond GRB pulses independently detectable by the Si-Tracker, MC and Super-AGILE.

- **Diffuse Galactic and extragalactic emission.** The AGILE good angular resolution and large average exposure will further improve our knowledge of cosmic ray origin, propagation, interaction and emission processes. We also note that a joint study of gamma-ray emission from MeV to TeV energies is possible by special programs involving AGILE and new-generation TeV observatories of improved angular resolution.

- **Gamma-ray pulsars.** AGILE will contribute to the study of gamma-ray pulsars in several ways: (1) improving photon statistics for gamma-ray period searches; (2) detecting possible secular fluctuations of the gamma-ray

Table 1: **AGILE Scientific Performance**

Gamma-ray Imaging Detector (GRID)	
Energy Range	30 MeV – 50 GeV
Field of view	~ 3 sr
Sensitivity at 100 MeV (ph cm ⁻² s ⁻¹ MeV ⁻¹ , 5 σ in 10 ⁶ s)	6 \times 10 ⁻⁹
Sensitivity at 1 GeV (ph cm ⁻² s ⁻¹ MeV ⁻¹ , 5 σ in 10 ⁶ s)	4 \times 10 ⁻¹¹
Angular Resolution at 1 GeV (68% cont. radius)	36 arcmin
Source Location Accuracy (S/N~10)	~5–20 arcmin
Energy Resolution (at 300 MeV)	$\Delta E/E \sim 1$
Absolute Time Resolution	~ 1 μ s
Deadtime	~ 100 μ s
Hard X-ray Imaging Detector (Super-AGILE)	
Energy Range	10 – 40 keV
Field of view (FW at Zero Sens.)	107 $^\circ$ \times 68 $^\circ$
Sensitivity (at 15 keV, 5 σ in 1 day)	~5 mCrab
Angular Resolution (pixel size)	~ 6 arcmin
Source Location Accuracy (S/N~10)	~2-3 arcmin
Energy Resolution	$\Delta E < 4$ keV
Absolute Time Resolution	~ 4 μ s
Deadtime (for each of the 16 readout units)	~ 4 μ s
Mini-Calorimeter	
Energy Range	0.3 – 200 MeV
Energy Resolution (above 1 MeV)	~ 1 MeV
Absolute Time Resolution	~ 3 μ s
Deadtime (for each of the 32 CsI bars)	~ 20 μ s

emission from neutron star magnetospheres; (3) studying unpulsed gamma-ray emission from plerions in supernova remnants and searching for time variability of pulsar wind/nebula interactions, e.g., as in the Crab nebula.

- **Search for non-blazar gamma-ray variable sources in the Galactic plane**, currently a new class of unidentified gamma-ray sources such as GRO J1838-04 ³³).

- **Galactic sources, new transients.** A large number of gamma-ray sources near the Galactic plane are unidentified, and sources such as 2CG 135+1 can be monitored on timescales of months/years. Also Galactic X-ray jet sources (such as Cyg X-3, GRS 1915+10, GRO J1655-40 and others) can pro-

duce detectable gamma-ray emission for favorable jet geometries, and a TOO program is planned to follow-up new discoveries of *micro-quasars*.

• **Fundamental Physics: Quantum Gravity.** The existence of sub-millisecond GRB pulses lasting hundreds of microseconds⁶⁾ opens the way to study QG delay propagation effects by AGILE detectors³⁶⁾. Particularly important is the AGILE Mini-Calorimeter with the independent readout for each of the 32 CsI bars of small deadtime ($\sim 20 \mu\text{s}$) and absolute timing resolution ($\sim 3 \mu\text{s}$). Energy dependent time delays near $\sim 100 \mu\text{s}$ for ultra-short GRB pulses in the energy range 0.3–3 MeV can be detected (requiring the detection of a minimum of ~ 5 photons). If these GRB ultra-short pulses originate at cosmological distances, sensitivity to the Planck's mass can be reached³⁶⁾.

References

1. Bakaldin A., Morselli A., Picozza P. et al., 1997, *Astroparticle Physics*, 8,109.
2. Barbiellini G. et al., 1995, *Nuclear Physics B*, 43, 253.
3. Barbiellini G. et al., 1995, *Nucl. Instrum. & Methods*, 354, 547.
4. Barbiellini, G., et.al., 2001, *Proceedings of the Symposium Gamma-2001*, AIP Conf. Proceedings, eds. S. Ritz, N. Gehrels, C.R. Shrader, Vol. 587, p. 774.
5. Barbiellini G., et al., al. , 2002, *NIM*, in press.
6. Bhat, C.L., et al., 1992, *Nature*, 359, 217.
7. Bignami, G.F. & Hermsen, W., 1993, *Ann. Rev. Astr. Astrophys.*, 21, 67.
8. Browning, R., Ramsden, D. & Wright, P.J., 1971, *Nature*, 232, 99.
9. D'Amico, N. et al., 2001, *Proceedings of the Symposium Gamma-2001*, AIP Conf. Proceedings, eds. S. Ritz, N. Gehrels, C.R. Shrader, Vol. 587, p. 555.
10. D'Amico, N. et al., 2001, *ApJL*, 552, L45.
11. De Jager, O.C., et al., 1996, *ApJ*, 457, 253.

12. Dingus, B.L., et al., 1994, in *Second Huntsville Gamma-Ray Burst Workshop*, eds. G.J. Fishman, J.J. Brainerd & Hurley, K., AIP Conf. Proceeding no. 307, p. 22.
13. Fermi, E., 1949, *Phys. Rev.*, 75, 1169.
14. Fermi, E., 1954, *ApJ*, 119, 1.
15. Fichtel, C.E. et al., 1975, *ApJ*, 198, 163.
16. Fichtel, C.E., 1994, *ApJS*, 94, 551.
17. Fichtel, C.E. & Trombka, J.I., 1997, *Gamma-Ray Astrophysics*, NASA Reference Publication No. 1386, Second Edition.
18. Golden R.L., Morselli A., Picozza P. et al., 1990, *Il Nuovo Cimento B*, 105, 191.
19. Hartman, R.C., Collmar, W., von Montigny, C. & Dermer, C.D., 1997, in *Proceedings of the Fourth Compton Symposium*, AIP Conf. Proceeding no. 401, p. 307.
20. Hartman R.C., et al., 1999, *ApJS*, 123, 279;
(<http://coss.gsfc.nasa.gov/coss/egret/3rd-EGRET-Cat.html>).
21. Hunter, S.D., et al., 1997, *ApJ*, 481, 205.
22. Hurley K. et al., 1994, *Nature*, 372, 652.
23. Kniffen, D.A. & Fichtel, C.E., 1970, *ApJ*, 161, L157.
24. Koyama, K., et al., 1995, *Nature*, 378, 255.
25. Kraushaar, W.L., et al., 1972, *ApJ*, 177, 341.
26. Prest M., et al., 2001, submitted to *NIM*.
27. Schneid E.J. et al., 1996a, in AIP Conf. Proc. no. 384, p.253.
28. Schoenfelder, V., et al., 1993, *ApJS*, 86, 657-692.
29. Sreekumar, P. & Fichtel, C.E., 1991, *A&A*, 251, 447.
30. Sreekumar, P., et al., 1993, *Phys. Rev. Letters*, 70, 127.

31. Sreekumar, P., et al., 1998, ApJ, 461, 872.
32. Swanenburg, B.N., et al., 1981, ApJ, 461, 872.
33. Tavani, M., et al., 1997, ApJ, 479, L109.
34. Tavani M., et.al., 2000, Proceedings of the 5th Compton Symposium, AIP Conf. Proceedings, ed. M. McConnell, Vol. 510, p. 746.
35. Tavani M., et.al., 2001, Proceedings of the Symposium *Gamma-2001*, AIP Conf. Proceedings, eds. S. Ritz, N. Gehrels, C.R. Shrader, Vol. 587, p. 729.
36. Tavani, M., et al., 2002, to be submitted.
37. Thompson D.J. et al., 1993, ApJS, 86, 629.
38. Thompson D.J., et al., 1995, ApJSS, 101, 259.
39. Thompson D.J., et al., 1996, ApJS, 107, 227.
40. Thompson D.J., Harding A.K., Hermsen W. & Ulmer M.P., 1998, in Proc. 4th CGRO Symp., eds. C.D. Dermer, M.S. Strickman, and J.D. Kurfess (New York, AIP Conf. Proc. no. 410), p. 39.
41. Tsyтович, V.N., 1966, Sov. Physics Usp., 9, 370.
42. Tsyтович, V.N., 1977, *Theory of Turbulent Plasma*, (New York: Plenum).

The GLAST Gamma Ray Large Area Telescope

Ronaldo Bellazzini, Nicola Omodei, Nick Lumb, Luca Baldini, Luca Latronico
INFN Pisa

ABSTRACT

The Gamma-ray Large Area Space Telescope (GLAST) is an international space mission that will study the cosmos in the energy range $10\text{ KeV} - 300\text{ GeV}$, the upper end of which is one of the last poorly observed regions of the celestial electromagnetic spectrum to be explored.

GLAST will have an imaging gamma-ray telescope vastly more capable than instruments flown previously. The main instrument, the Large Area Telescope (LAT), will have superior area, angular resolution, field of view, and dead time that together will provide a factor of 60 or more advance in sensitivity, and capability for the study of transient phenomena.

1 Introduction

One of the last bands of the electromagnetic spectrum to be explored for astronomy is the range above 20 MeV . The principal reason for the late start

was technological: for energies up to tens of GeV , detectors must be placed in orbit, and even from orbit detection of the low fluxes of celestial gamma rays is difficult.

First came EGRET (Launched in 1991): it made the first complete survey of the sky in the $30 MeV - 10 GeV$ range. The main discoveries of this mission were the detection of gammas with energy $> 100 MeV$ coming from Active Galactic Nuclei (observed more than 60) and the measurement of diffuse gamma ray background to over $10 GeV$. But the majority of the sources that shine in the gamma sky don't have a counterpart in low energy: one hundred and seventy sources in the 3rd EGRET catalog are unidentified. GLAST will enable identification of EGRET sources by providing much finer positional error bounds.

EGRET raised many interesting issues and questions which can be addressed by a NASA mid-class mission (Delta II rocket). The GLAST mission was conceived to address important outstanding questions in high-energy astrophysics, many of which were raised but not answered by results from EGRET. The main instrument on board the GLAST detector is the Large Area Telescope (LAT) that is a pair conversion telescope, like EGRET, but the detectors will be based on solid-state technology, obviating the need for consumables and greatly decreasing instrument deadtime. In this paper we will describe the development of the LAT detector and we will focus our attention in some scientific topics of interest for GLAST.

2 The Large Area Telescope

The primary interaction of photons with matter in the GLAST energy range is pair conversion. This process forms the basis for the underlying measurement principle by providing an unique signature for gamma rays, which distinguish them from charged particles. The flux of Cosmic rays, in fact, is as much as 10^5 times larger. The pair conversion process permits the determination of the incident photon directions via the reconstruction of the trajectories of the resulting e^+e^- pairs. This technique is illustrated in Figure 1 in which the incident radiation first passes through an anticoincidence shield, which is sensitive to charged particles, then through thin layers of high-Z (tungsten) material called *conversion foils*. The photon converts in these layers producing an electron-positron pair. The trajectories of these charged particles are

measured by the tracking detectors, and their energies are then measured by a calorimeter. *GLAST* was designed to have a low profile to give wide field of view.

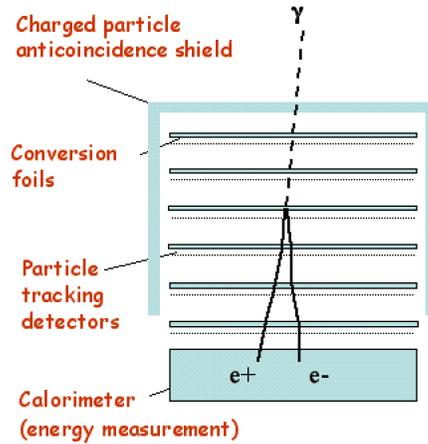


Figure 1: *Principle of photon detection in GLAST.*

The Large Area Telescope (LAT) comprises an array of 16 identical “tower” modules (see Figure 2), each with a tracker (Si strips) and a calorimeter (CsI with PIN diode readout) and DAQ module. The towers are surrounded by a finely segmented ACD (plastic scintillator with PMT readout) while the support structure is an aluminum strong-back “Grid” with heat pipes for transport of heat to the instrument sides.

The **Anticoincidence Detector** (ACD) has a segmented plastic scintillator to minimize self-veto at high Energy and to enhance the background rejection: the estimated efficiency is greater than 0.9997. The purpose of the ACD is to detect incident charged cosmic ray particles that outnumber cosmic gamma rays by more than 5 orders of magnitude. Signals from the ACD can be used as a trigger veto or can be used later in the data analysis ²⁾.

Each of the 16 **Tracker** tower modules consists of a stack of 19 “tray” structures. Silicon detector wafers cover either side of a tray with the strips on each side running in the same direction. Every other tray is rotated by 90°, so

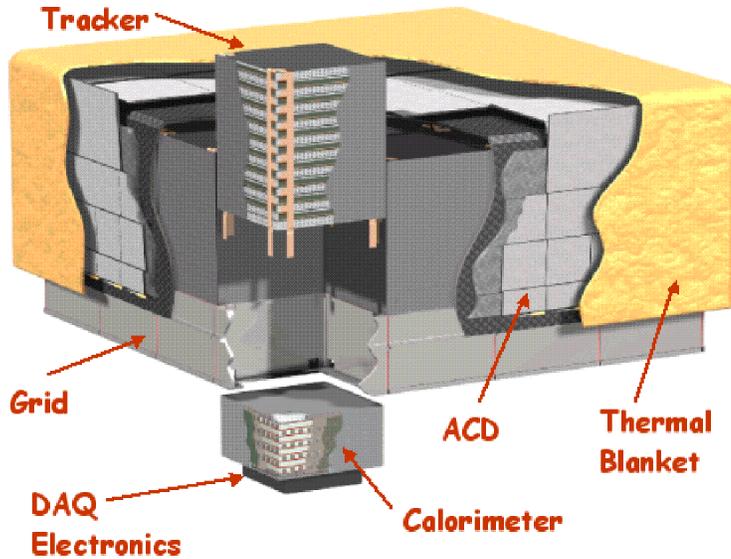


Figure 2: *The LAT instrument components*

each W foil is followed immediately by an x, y plane of detectors with a $2mm$ gap between x and y layers. The detectors are located close to the conversion foils to minimize multiple-scattering errors. The bottom tray has a flange to mount on the support grid. The electronics hybrids are glued vertically to the tray sides to minimize the gap between towers. Each silicon plane on a tray has a $37cm \times 37cm$ active cross section, giving a total silicon area of $83m^2$ (comparable with the ATLAS detector planned for the CERN LHC project). In all there are 11500 Silicon Strip Detectors and a total of 1 million channels.

The **Calorimeter** is made of $96CsI$ crystals (thallium doped) per tower arranged into a hodoscopic, imaging configuration and with PIN diode read-out on each end. The electronics chain for each PIN diode is composed of a preamplifier which feeds two shaping amplifiers. Discriminators divide the

energy domain into four energy ranges, two peak-detecting track and holds. A third faster shaping amplifier, peaking at $0.5\mu s$ is used for fast trigger discrimination. The main features of the calorimeter detector are the large dynamic range (5×10^5), low nonlinearity (less than 2%), low power consumption, and minimal dead time (less than $20\mu s$ per event).

The LAT trigger is a 3-level system. Primary requirements are high efficiency for all measurable gamma rays, and background reduction to fit with telemetry capacity. Two separate conditions may initiate a hardware trigger for a given tower (LT1). The first request is for the tracker to have three planes hit in a row. The second involves the calorimeter, considering the number of hits in the module. Tower triggers are ORd in the central ACD-TEM and fanned out to each tower. The ACD information is optionally used to reduce LT1 rate (“controlled mode”). The second level trigger (LT2) is a tower-based trigger, in parallel for all towers. It uses a fast track finding algorithm and extrapolates track candidates to the ACD tiles to search for vetoes. The veto is not applied to events with large energy deposits in CAL. LT3 is a full instrument event reconstruction trigger. The main features of the three level trigger are summarised in Table 1. Albedo photon events are removed by comparing the reconstructed photon direction with that of the Earth’s horizon. The cosmic ray event rate is reduced to less than 15 Hz.

Level	Type	Location	Components	Function	Pk Rate	Avg. Rate
L1: initiate readout of the detectors.	Hardware	OR of independent triggers in each tower.	TKR: coinc. of x,y planes CAL-LOW: # of hits CAL-HIGH: energy ACD: high threshold	Two redundant triggers for gamma-rays. Avoid self-veto at high E. Select C,N,O for calibration.	9 kHz (3.4 kHz with ACD veto enabled)	5.5 kHz (2 kHz with ACD veto enabled)
L2: cosmic-ray rejection	Software	Individual towers + ACD	L1 information. Simple track reconstruction. Extrapolation to ACD.	Reject tracks that point to fired ACD tiles, unless CAL energy is high.	1.7 kHz	1 kHz
L3: final on-board background rejection	Software	SIU (Full Instrument)	Full event reconstruction (all subsystems). SC ancillary data (attitude information).	Loose cuts to reject background, including Earth albedo, sufficiently for downlink.		<30 Hz

a. Rates are calculated from detailed simulations of the backgrounds, the detector response, and the trigger logic.

Table 1: *The 3 different trigger levels adopted in the GLAST Large Area Telescope.*

3 Summary of GLAST Science Topics

The universe is largely transparent to gamma rays in the energy range of GLAST. Energetic sources near the edge of the visible universe can be detected by the light of their gamma rays. There are good reason to expect that GLAST will see known classes of sources up to redshift 5, or even greater if the sources existed at earlier times. The small interaction cross section for gamma rays can provide a direct view into nature's highest energy accelerators. In addition, gamma rays point back to their sources unlike cosmic rays which are deflected by magnetic fields.

The main advantages of the LAT detector will be the wide field of view ($2sr$) and the extremely short dead time per event ($< 100\mu s$). These performances, together with the excellent background rejection (better than $2.5 \times 10^5 : 1$) will allow GLAST to detect both faint sources and transient signals in the gamma-ray sky. The capabilities of the GLAST LAT detector compared to those of *EGRET* are summarized in table 2. Several performances of the LAT detector, such as the angular and energy resolution, the field of view and the effective area are plotted in figure 3, and compared to those of EGRET.

Quantity	EGRET	LAT (Minimum Spec.)
Energy Range	20 MeV–30 GeV	20 MeV–300 GeV
Peak Area	1500 cm ²	8000 cm ²
Field of View	0.5 sr	> 2sr
Angular Resolution	5.8°	< 3.5° (100 MeV) < 0.15° (> 10 GeV)
Energy resolution	10%	10%
Deadtime per event	100 ms	< 100μs
Source Location Determination	15'	< 0.5'
Point Source Sensitivity	$1 \times 10^{-7} \text{cm}^{-2} \text{s}^{-1}$	$< 6 \times 10^{-9} \text{cm}^{-2} \text{s}^{-1}$

Table 2: *GLAST LAT specification and performance compared with EGRET*

EGRET discovered that a class of **Active Galactic Nuclei** (AGN), known as blazars, is a bright and variable source of high energy gamma rays (for reference see the 3rd EGRET catalog reported in figure 4). The peak in energy from many blazars is emitted in the GLAST energy band. The emission is believed to be powered by accretion onto supermassive black holes

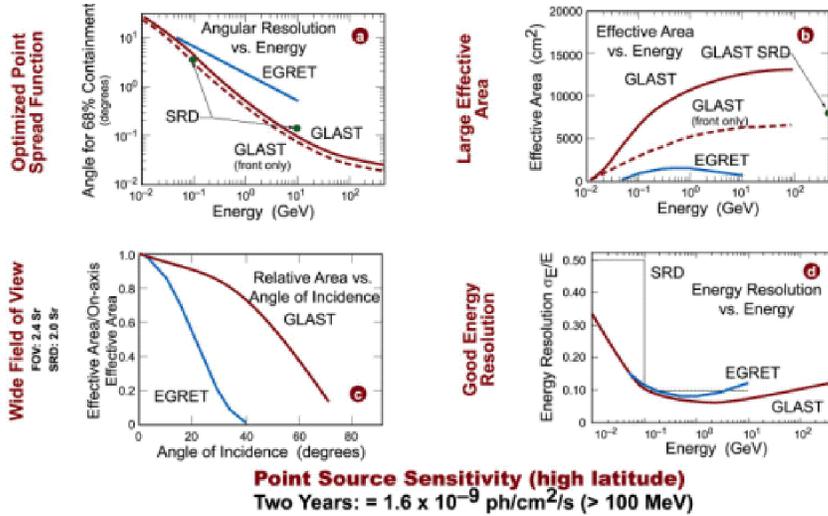


Figure 3: *LAT* detector performance compared with *EGRET* for a point source observation.

in the centre of distant galaxies. *GLAST* will be able to extend the number of AGN gamma-ray sources from 70 to thousands: it will be an all-sky monitor for AGN flares scanning the whole sky every three hours. Moreover it will be able to study the variability of the flares decreasing the minimum time scale for detection of variability.

Due to its short dead time *GLAST* will continue the recent revolution of **Gamma Ray Bursts** (GRB), measuring the variability of signals at high energy and tracking the gamma ray afterglow ¹). There are good reasons to think that with the information collected by *GLAST* it will be possible to discriminate between several emission models, starting for example from the signatures that different theoretical scenarios leave in the signal; the presence of internal shocks or the emission of high energy photons due to the collision

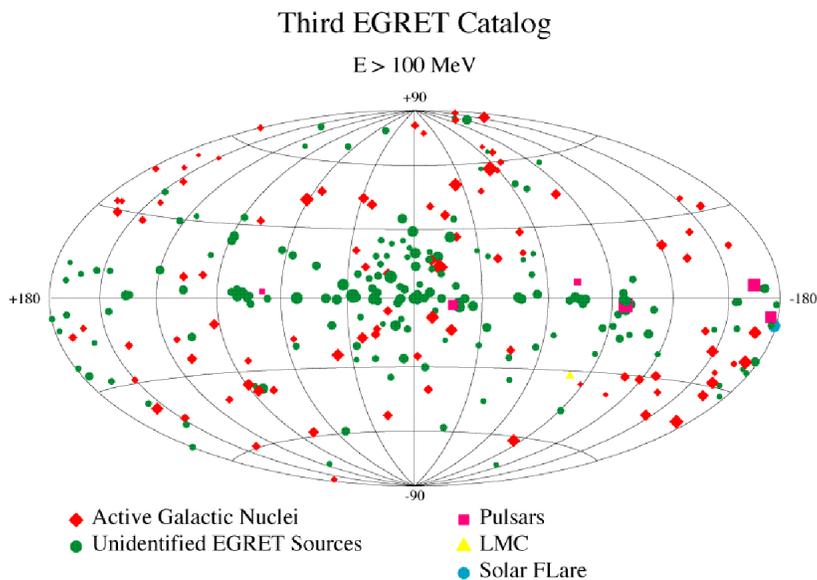


Figure 4: *3rd EGRET catalog for sources of energies greater than 100 MeV. The majority of the sources result still unidentified.*

between a single shell and the surrounding medium are some examples.

Another breakthrough could be the direct discovery of **Dark Matter Halos**, and the understanding of the nature of the matter from which they are composed ³⁾. The large area and the low instrumental background of GLAST will also allow searches for the decay of exotic particles in the early universe and for annihilations of postulated weakly-interacting massive particles (or *WIMPs*). Most of the isotropic background measured by EGRET will be resolved by GLAST as diffuse emissions from AGN. A diffuse cosmic residual would be a tremendous discovery and could be related to strange particle decay. Recent work suggests that annihilation of the lightest supersymmetric particles, a candidate Galactic Halo WIMP, could leave a signature in the

diffuse background detectable by GLAST.

GLAST will also be able to discovery many pulsars that emit in the gamma-ray energy band. The number of pulsars detected during the GLAST lifetime will be potentially 250 or more, providing good statistics to distinguish between the two primary models proposed to describe the particle acceleration and gamma ray production: the outer gap and the polar cap models.

Equally important, GLAST will spatially resolve the remnants of supernovae explosions (SNR), precisely measuring their spectra. In principle these observations could be address the problem of whether SNRs are the source of cosmic rays. Cosmic rays are the main sources of production of the gamma diffuse background in the Milky Way via their interaction with the interstellar medium. GLAST will detect the diffuse emission from a number of local group galaxies and map their emission within the largest of these for the first time. Spatial and spectral studies will test cosmic-ray production and diffusion models ^{4, 5}).

4 Conclusion

GLAST will provide an important step forward in gamma astronomy and in the comprehension of the nature of the high energy universe. GLAST is planned as a facility-class mission involving an international collaboration from the particles physics and astrophysics communities. Currently scientists from the United States, Japan, France, Germany and Italy are involved in detector construction, testing, assembling and in software development. GLAST is planned to be launched early in 2006 from the Kennedy Space Center on board a Delta 2920 vehicle. It will have a 550 km altitude circular orbit, with an inclination of 28.5° .

For more information see the GLAST web site at:

<http://www-glast.stanford.edu/>

For information about the local Pisa activity see:

<http://www.pi.infn.it/glast/>

References

1. J.P. Norris *et al*, astro-ph/9912136 (1999)
2. A. Moiseev *et al*, astro-ph/9912138 (1999)
3. A. Moiseev *et al*, astro-ph/9912139 (1999)
4. J. F. Ormes *et al*, astro-ph/0003270 (2000)
5. S.W. Digel *et al*, astro-ph/0003407 (2000)
6. F.W. Stecker *et al*, astro-ph/9909157 (1999)
7. T. Kamae *et al*, astro-ph/9901187 (1999)

Frascati Physics Series Vol. XXIV (2002), pp. 363-380
INTERNATIONAL SCHOOL OF SPACE SCIENCE - 2001 Course on
ASTROPARTICLE AND GAMMA-RAY PHYSICS IN SPACE, L'Aquila (Italy), 30/8 - 7/9, 2001

GAMMA RAY ASTROPARTICLE PHYSICS WITH GLAST

Aldo Morselli

Dept. of Physics, Univ. of Roma "Tor Vergata" and INFN Roma 2, Italy

ABSTRACT

The energy domain between 10 MeV and hundreds of GeV is an essential one for the multifrequency study of extreme astrophysical sources. The understanding of spectra of detected gamma rays is necessary for developing models for acceleration, emission, absorption and propagation of very high energy particles at their sources and in space. After the end of EGRET on board the Compton Gamma Ray Observatory this energy region is not covered by any other experiment, at least up to 50 GeV where ground Cerenkov telescopes are beginning to take data. Here we will review the status of the space experiment GLAST that will fill this energy region from March 2006 with particular emphasis at the connection with all the other ground and space planned experiments and at the contribution of GLAST to particle physics.

1 GLAST

The techniques for the detection of gamma-rays in the pair production regime energy range are very different from the X-ray detection ones. For X-rays detection focusing is possible and this permits large effective area, excellent energy resolution, very low background. For gamma-rays no focusing is possible and this means limited effective area, moderate energy resolution, high background but a wide field of view (see figure 1). This possibility to have a wide field of view is enhanced now, in respect to EGRET, with the use of silicon detectors, that allow a further increase of the ratio between height and width (see fig.2), essentially for two reasons: a) an increase of the position resolution that allow a decrease of the distance between the planes of the tracker without affect the angular resolution, b) the possibility to use the silicon detectors themselves for the trigger of an events, with the elimination of the Time of Flight system, that require some height.

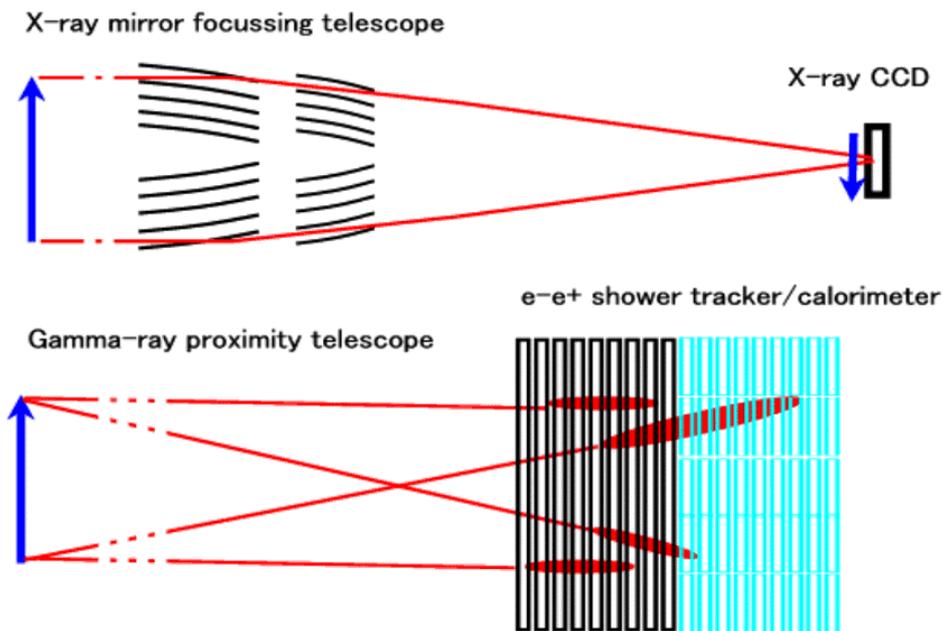


Figure 1: *Detector Technology: X-ray versus Gamma-ray.*

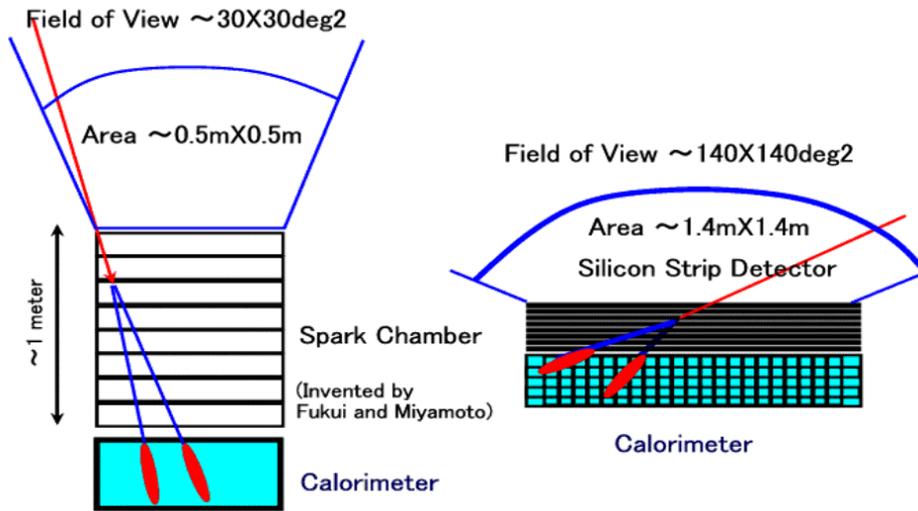


Figure 2: *EGRET* (Spark Chamber) versus *GLAST* (Silicon Strip Detector).

The Gamma-ray Large Area Space Telescope (*GLAST*)¹⁾, has been selected by NASA as a mission involving an international collaboration of particle physics and astrophysics communities from the United States, Italy, Japan, France and Germany for a launch in the first half of 2006. The main scientific objects are the study of all gamma ray sources such as blazars, gamma-ray bursts, supernova remnants, pulsars, diffuse radiation, and unidentified high-energy sources. Many years of refinement has led to the configuration of the apparatus shown in figure 3, where one can see the 4x4 array of identical towers each formed by:

- Si-strip Tracker Detectors and converters arranged in 18 XY tracking planes for the measurement of the photon direction.
- Segmented array of CsI(Tl) crystals for the measurement the photon energy.
- Segmented Anticoincidence Detector (ACD).

The main characteristics are an energy range between 20 MeV and 300 GeV, a field of view of ~ 3 sr, an energy resolution of $\sim 5\%$ at 1 GeV, a point source sensitivity of 2×10^{-9} (ph cm⁻² s⁻¹) at 0.1 GeV, an event deadtime of 20 μ s and a peak effective area of 10000 cm², for a required power of 600 W and a payload weight of 3000 Kg.

The list of the people and the Institution involved in the collaboration together with the on-line status of the project is available at <http://www-glast.stanford.edu>. A description of the apparatus can be found in ²⁾.

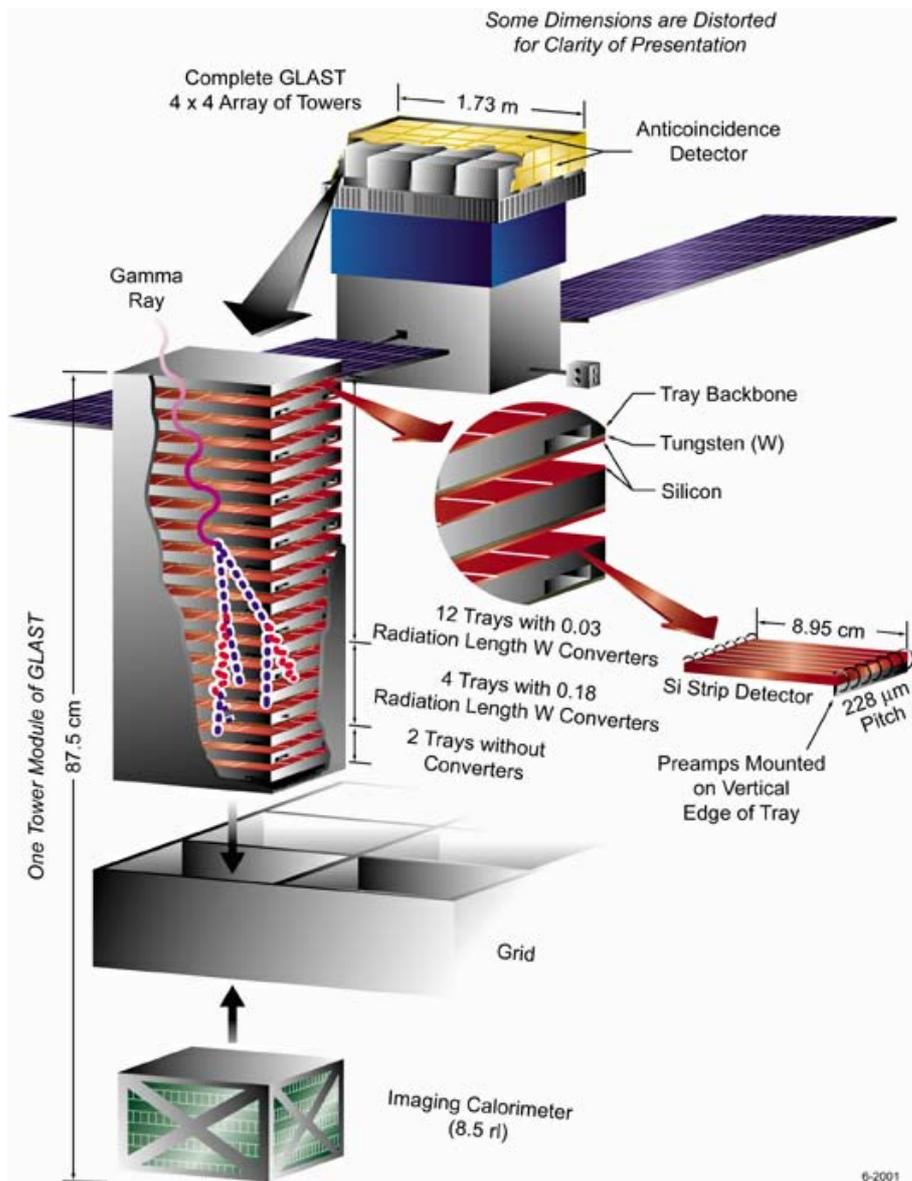


Figure 3: *The GLAST instrument, exploded to show the detector layers in a tower, the stacking of the CsI logs in the calorimeter, and the integration of the subsystems.*

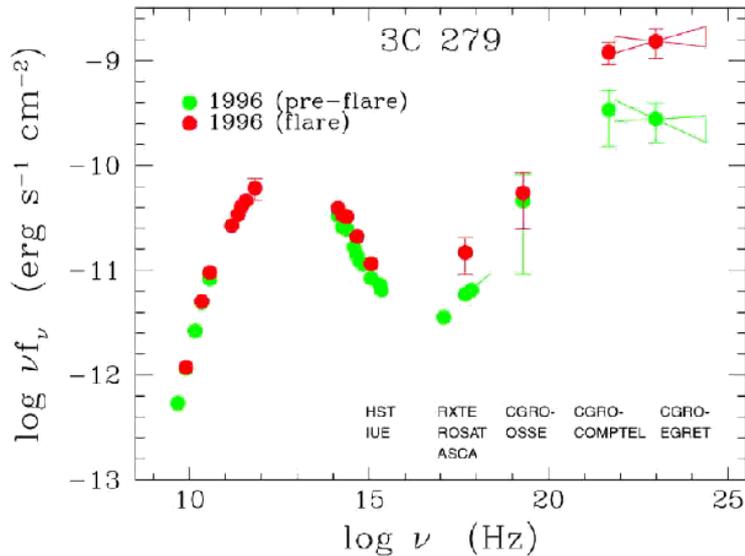


Figure 4: *Spectral energy distributions of the quasars 3C 279 during flaring state (in red) and non flaring state (in green).*

1.1 Active Galactic Nuclei

Before EGRET, 3C 273 was the only active galactic nucleus (AGN) known to emit high-energy gamma rays. Now we know that there is an entire class of active galaxies that probably represent the largest class of high energy gamma-ray emitters: the blazars. Blazars are flat radio spectrum, active galactic nuclei, or AGN, whose members include BL Lac objects and highly polarized and optically violently variable quasars that often emits more in gamma-ray than in any other frequencies (see figure 4). For a review on AGNs see reference ³). GLAST will dramatically extend the number of observed AGNs, as well as the energy range over which they can be observed. Indeed, GLAST might be called the "Hubble Telescope" of gamma-ray astronomy as it will be able to observe AGN sources to $z \sim 4$ and beyond, if such objects actually existed at such early times in the universe. Figure 5 shows the so called Log N versus Log S distribution, where N is the number of sources and S is source flux for $E_\gamma > 100$ MeV, for AGN. The curve is extrapolated from EGRET data and an AGN model of the diffuse gamma-ray background based on the assumption that

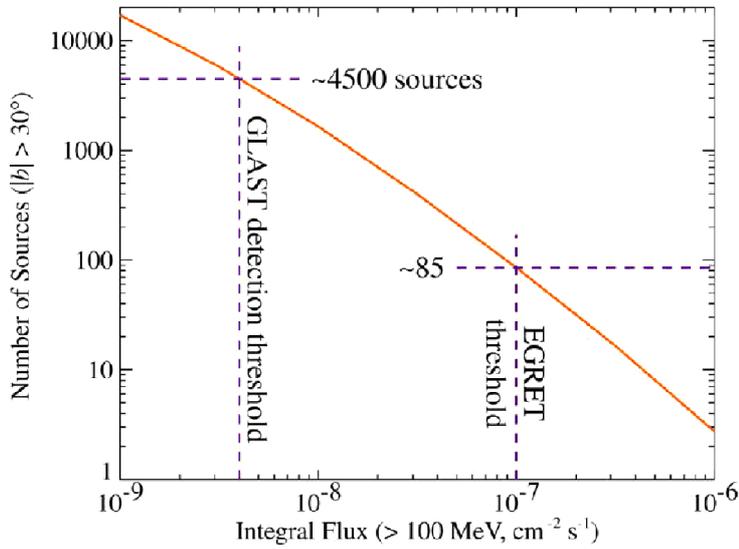


Figure 5: Estimate of the number of AGNs that GLAST will detect at high latitude in a 2 year sky survey compared to EGRET's approximate detection limit.

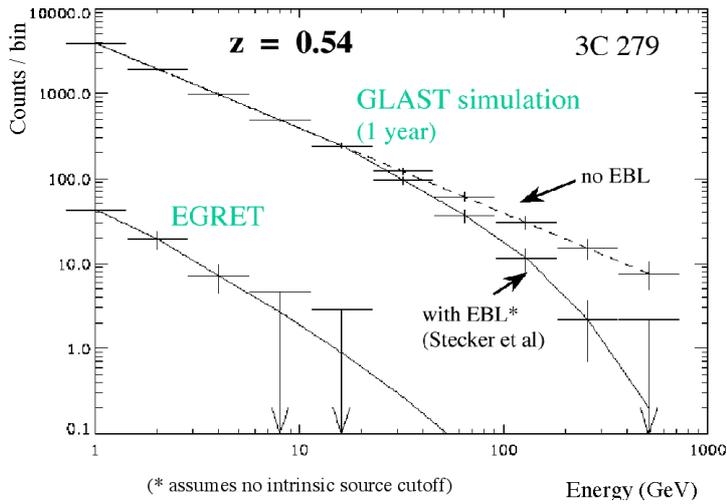


Figure 6: Number of photons detected by EGRET from 3C279 and the number expected with GLAST in the case of extragalactic background light attenuation and without attenuation.

AGN sources follow a luminosity function similar to flat spectrum radio quasars. Extrapolation from EGRET AGN detections projects that about 5,000 AGN sources will be detected in a 2 year cumulative scanning mode observation by GLAST, as compared to the 85 that have been observed by EGRET in a similar time interval. This large number of AGN's covering a redshift range from $z \sim 0.03$ up to $z \sim 4$ will allow to disentangle an intrinsic cutoff effect, i.e., intrinsic to the source, from a cut-off derived from the interaction with the extra galactic background light, or EBL. Only by observing many examples of AGN, and over a wide range of redshifts, one can hope to untangle these two possible sources of cutoff. In figure 6 is shown the number of photons detected by EGRET from 3C279 and the number expected with GLAST in the case of extragalactic background light attenuation and without attenuation. Determination of the EBL can provide unique information on the formation of galaxies at early epochs, and will test models for structure formation in the Universe.

1.2 Gamma-ray burst

Gamma-ray bursts (GRBs) are intermittently the most intense and most distant known sources of high-energy gamma rays; at GeV energies, the brightest GRBs are 1000-10,000 times brighter than the brightest AGN. The unparalleled luminosities and cosmic distances of GRBs, combined with their extremely fast temporal variability, make GRBs an extremely powerful tool for probing fundamental physical processes and cosmic history.

GLAST, in concert with the Gamma-ray Burst Monitor, will measure the energy spectra of GRBs from a few keV to hundreds of GeV during the short time after onset when the vast majority of the energy is released. GLAST will also promptly alert other observers, thus allowing the observations of GLAST to be placed in the context of multiwavelength afterglow observations, which are the focus of HETE-2 and the upcoming Swift missions. The additional information available from GLAST's spectral variability observations will be key to understanding the central engine.

Figure 7 illustrates a very intense, short GRB. The true EGRET time profile is very uncertain because the \sim two hundred milliseconds EGRET dead time per photon is comparable to GRB pulse widths; hence, many more photons may have been incident on EGRET during the extremely intense initial pulse. The GLAST dead time will be \sim 10,000 times smaller, thus allowing

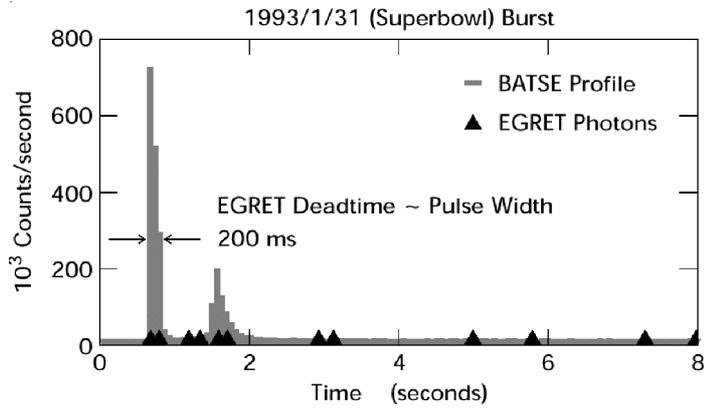


Figure 7: *EGRET* and *BATSE* light curves of the Superbowl burst, *GRB930131*. The burst consisted of an extremely intense spike, followed by low-level emission for several seconds. The true temporal development at energies >100 MeV is uncertain since *EGRET* dead time is comparable to *GRB* pulse widths.

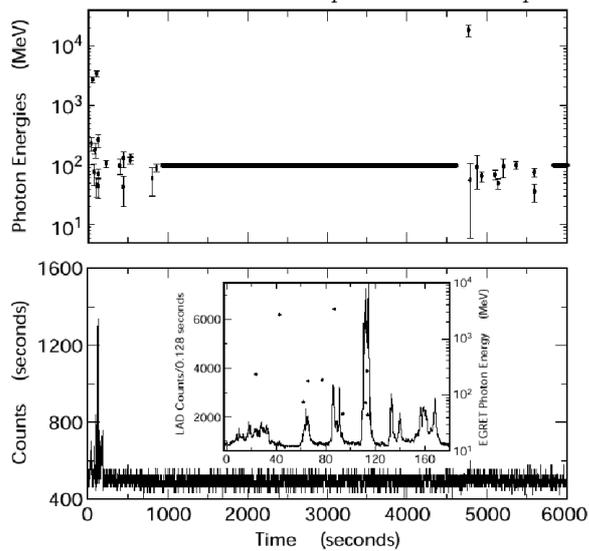


Figure 8: *EGRET* and *BATSE* light curves of *GRB940217*. Burst cessation at *BATSE* energies occurs at 160 s. Extended emission at *EGRET* energies persist beyond an intervening earth occultation, up to 5000 seconds after the *BATSE* event.

a precise measurement of the gamma-ray flux during the peak. This characteristic together with its larger field of view and larger effective area, should permit to detect virtually all GRBs in its field of view reaching the "the edge" of the GRB distribution, as does BATSE. Figure 8 shows another intense burst with very different temporal character which occurred in EGRET's field of view on 1994 Feb 17. At BATSE energies (25 - 1000 keV), this event persisted for ~ 160 s; however, at EGRET energies, it apparently continued at a relatively high flux level past an Earth occultation, for at least 5000 s, to deliver a delayed ~ 18 GeV photon. GLAST, with negligible self veto, will have good efficiency above 10 GeV and it will be able to localize GRBs with sufficiently high accuracy to enable rapid searches at all longer wavelengths. About half of the 200 bursts per year detected by GLAST will be localized to better than 10 arc minute radius, an easily imaged field for large-aperture optical telescopes.

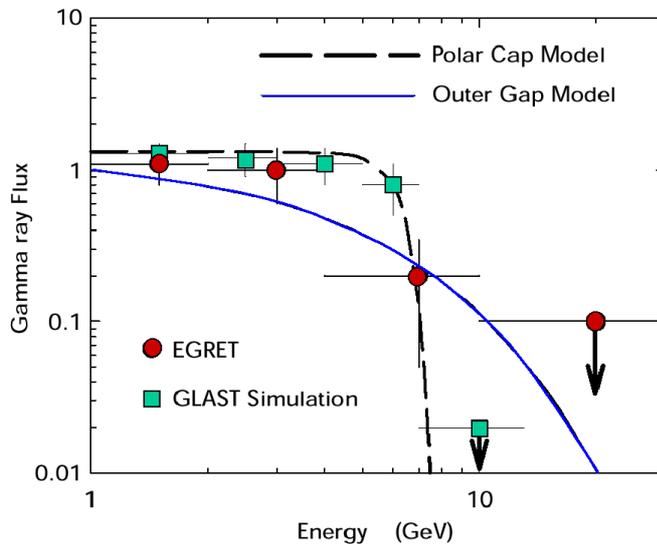


Figure 9: *Modeled high-energy pulsar spectrum, showing the improvement in resolution between EGRET and GLAST. The polar cap model predicts a sharp high-energy cutoff, while the outer gap model predicts a more gradual cutoff. Unlike EGRET, GLAST will be able to distinguish the true shape of the spectrum (assumed to be that of the polar cap model in this simulation).*

1.3 Pulsars

GLAST will discover many gamma-ray pulsars, potentially 50 or more, and will provide definitive spectral measurements that will distinguish between the two primary models proposed to explain particle acceleration and gamma-ray generation: the outer gap ⁴⁾ and polar cap models ⁵⁾ (see figure 9). From observations made with gamma ray experiments through the EGRET era, seven gamma-ray pulsars are known. GLAST will detect more than 100 pulsars and will be able to directly search for periodicities in all EGRET unidentified sources. Because the gamma-ray beams of pulsars are apparently broader than their radio beams, many radio-quiet, Geminga-like pulsars likely remain to be discovered.

1.4 Search for supersymmetric dark matter

GLAST is particularly interesting for the supersymmetric particle search because, if neutralinos make up the dark matter of our galaxy, they would have non-relativistic velocities, hence the neutralino annihilation into the gamma gamma and gamma Z final states can give rise to gamma rays with unique energies $E_\gamma = M_\chi$ and $E'_\gamma = M_\chi (1 - m_z^2/4M_\chi^2)$.

In figure 10 is shown how strong can be the signal ⁶⁾ in the case of a cuspy dark matter halo profiles distribution ⁷⁾.

Figure 11 shows the GLAST capability to probe the supersymmetric dark matter hypothesis ⁶⁾. The various zone sample the MSSM with different values of the parameters space for three classes of neutralinos. The previous galaxy dark matter halo profile ⁷⁾ that gives the maximal flux has been assumed. The solid line shows the number of events needed to obtain a 5σ detection over the galactic diffuse γ -ray background as estimated from EGRET data. As the figures show, a significant portion of the MSSM phase space is explored, particularly for the higgsino-like neutralino case.

This effort will be complementary to a similar search for neutralinos looking with cosmic-ray experiments like the next space experiment PAMELA ⁸⁾ at the distortion of the secondary positron fraction and secondary antiproton flux induced by a signal from a heavy neutralino.

In figure 12 (on the left) there are the experimental data ⁹⁾ for the positron fraction together with the distortion of the secondary positron fraction

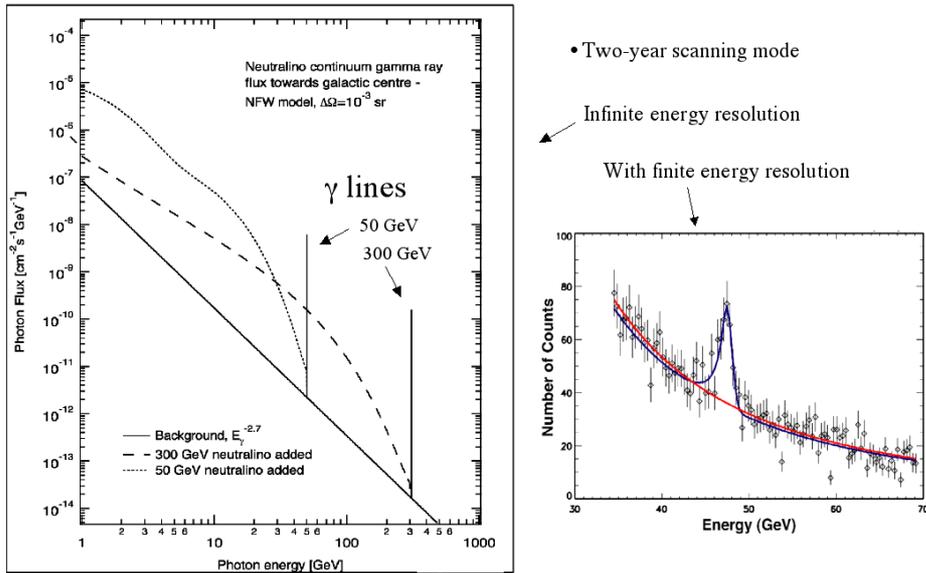


Figure 10: Total photon spectrum from the galactic center from $\chi\chi$ annihilation (on the left), and number of photons expected in GLAST for $\chi\chi \rightarrow \gamma\gamma$ from a 1-sr cone near the galactic center with a 1.5% energy resolution (on the right)

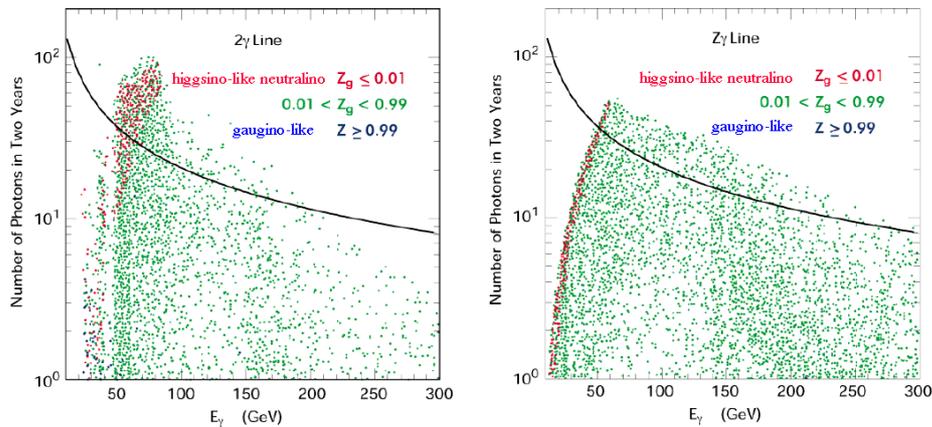


Figure 11: Number of photons expected in GLAST for $\chi\chi \rightarrow \gamma\gamma$ from a 1-sr cone near the galactic center as a function of the possible neutralino mass. The solid line shows the number of events needed to obtain a five sigma signal detection over the galactic diffuse gamma-ray background as estimated by EGRET data.

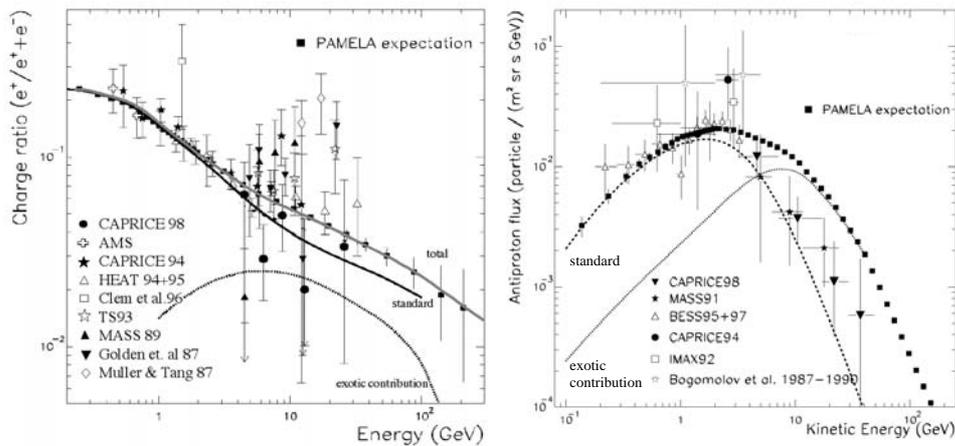


Figure 12: *Distortion of the secondary positron fraction (on the left) and secondary antiproton flux (on the right) induced by a signal from a heavy neutralino. The PAMELA expectation in the case of exotic contributions are shown by black squares*

(solid line) due to one possible contribution from neutralino annihilation (dotted line, from ¹⁰). The expected data from the experiment PAMELA in the annihilation scenario for one year of operation are shown by black squares ¹¹).

In the same figure (on the right) there are the experimental data for the antiproton flux ¹² together with the distortion on the antiproton flux (dashed line) due to one possible contribution from neutralino annihilation (dotted line, from ¹³). The antiproton data that PAMELA would obtain in a single year of observation for one of the Higgsino annihilation models are shown by black squares.

2 Conclusion

The gamma-ray space experiment GLAST is under construction. Its time of operation and energy range is shown together with the other space X-ray satellite and gamma-ray experiments in figure 13. Note that it will cover an interval not covered by any other experiments. Note also the number of other experiments in other frequencies that will allow extensive multifrequency studies.

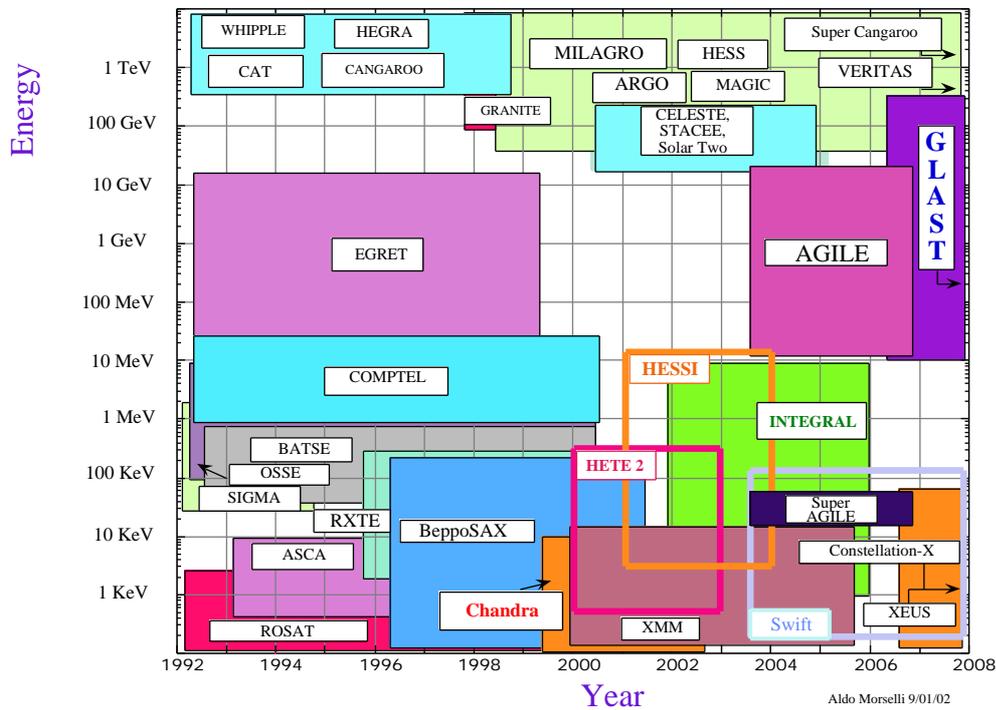


Figure 13: *Timeline schedule versus the energy range covered by present and future detectors in X and gamma-ray astrophysics.*

In the last decade, ground-based instruments have made great progress, both in technical and scientific terms. High-energy gamma rays can be observed from the ground by experiments that detect the air showers produced in the upper atmosphere. Air shower arrays directly detect the particles (electrons, muons, and photons) in air showers, and atmospheric Cherenkov telescopes detect the Cherenkov radiation created in the atmosphere and beamed to the ground. Detectors based on the atmospheric Cherenkov technique consist of one or more mirrors that concentrate the Cherenkov photons onto fast optical detectors. Photomultiplier tubes (PMTs) placed in the focal plane are generally used to detect the Cherenkov photons. Two problems in using atmospheric Cherenkov telescopes (ACT) are the night-sky background and the large isotropic background from cosmic-ray showers.

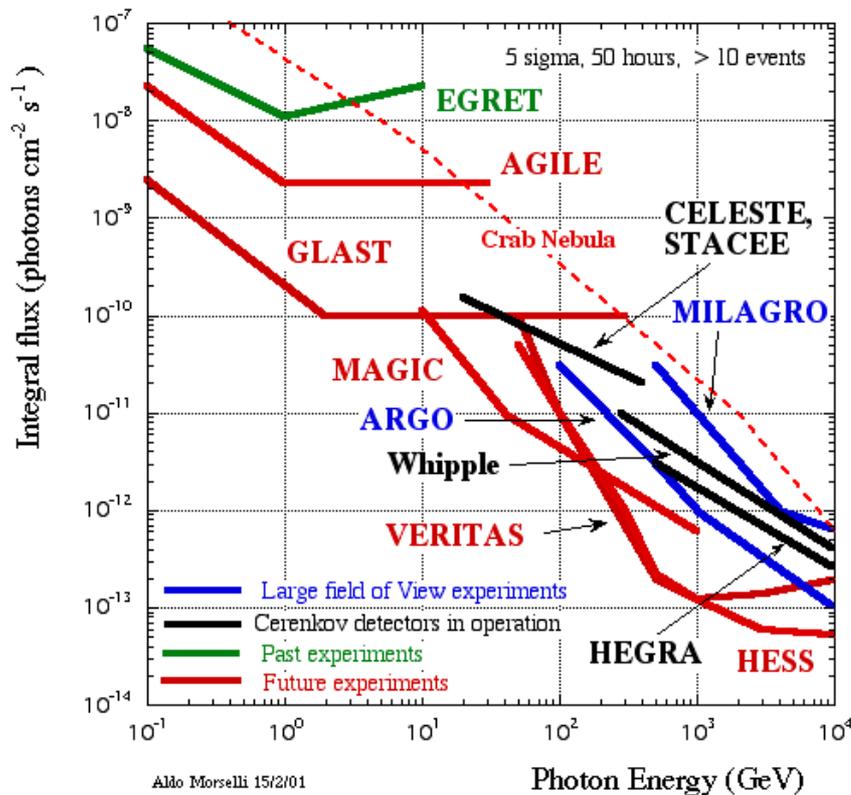


Figure 14: *Sensitivity of present and future detectors in the gamma-ray astrophysics.*

The energy threshold of an atmospheric Cherenkov telescope is determined by the number of Cherenkov photons needed to observe a signal above the level of the night-sky background. For individual point sources, ground-based instruments have unparalleled sensitivity at very high energies (above 50-250 GeV). For many objects, full multi-wave-length coverage over as wide an energy range as possible will be needed to understand the acceleration and gamma-ray production mechanisms. On the technical side, atmospheric Cherenkov telescopes have demonstrated that a high degree of gamma/hadron discrimination and a source pointing accuracy of 10-30 arc minutes (depending

on the source strength) can be achieved based on the detected Cherenkov image. Also the energy threshold is lowering remarkably (for a review, see ¹⁴). In figure 14 the GLAST sensitivity compared with the others present and future detectors in the gamma-ray astrophysics range is shown. The predicted sensitivity of a number of operational and proposed Ground based Cherenkov telescopes, CELESTE, STACEE, VERITAS, Whipple is for a 50 hour exposure on a single source. EGRET, GLAST, MILAGRO, ARGO and AGILE sensitivity is shown for one year of all sky survey. The diffuse background assumed is $2 \cdot 10^{-5} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} (100 \text{ MeV}/E)^{1.1}$, typical of the background seen by EGRET at high galactic latitudes. The source differential photon number spectrum is assumed to have a power law index of -2, typical of many of the sources observed by EGRET and the sensitivity is based on the requirement that the number of source photons detected is at least 5 sigma above the background. Note that on ground only MILAGRO and ARGO will observe more than one source simultaneously. The Home Pages of the various instruments are at <http://www-hfm.mpi-hd.mpg.de/CosmicRay/CosmicRaySites.html>.

A wide variety of experiments provide interesting probes for the search of supersymmetric dark matter. Indirect dark matter searches and traditional particle searches are highly complementary. In the next five years, an array of experiments will be sensitive to the various potential neutralino annihilation products. These include under-ice and underwater neutrino telescopes, atmospheric Cerenkov telescopes and the already described space detectors GLAST and PAMELA together with AMS. In many cases, these experiments will improve current sensitivities by several orders of magnitude. Direct dark matter probes share features with both traditional and indirect searches, and have sensitivity in both regions. In the cosmologically preferred regions of parameter space with $0.1 < \Omega_\chi h^2 < 0.3$, all models with charginos or sleptons lighter than 300 GeV will produce observable signals in at least one experiment. An example ¹⁵) is shown in figure 15 in the framework of minimal supergravity, which is fully specified by the five parameters (four continuous, one binary) $m_0, M_{1/2}, A_0, \tan \beta, \text{sgn}(\mu)$. Here, m_0 , $M_{1/2}$, and A_0 are the universal scalar mass, gaugino mass, and trilinear scalar coupling ¹⁵). The figure shows the limits that can be obtained in the $m_0, M_{1/2}$ plane for $\tan \beta = 10$, $A_0 = 0$, $\mu > 0$. Higher values (~ 50) of $\tan \beta$ requires significant fine-tuning of the electroweak scale. The limit from gamma-ray assumes a moderate halo profile.

Estimated reaches before LHC

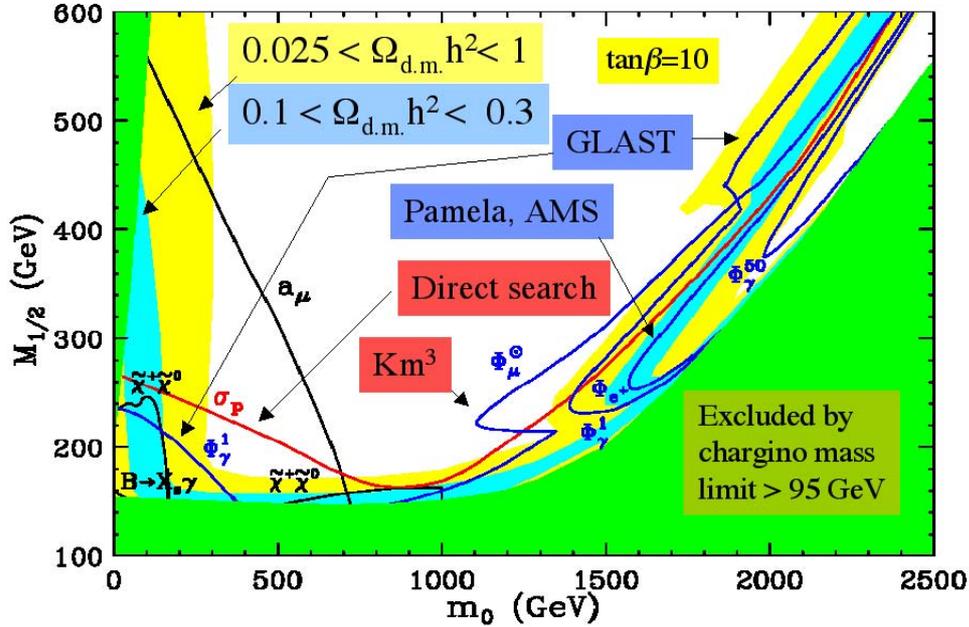


Figure 15: *Example of estimated reaches of various searches before the LHC begins operation. Note the complementarity between the different techniques. For moderate values of $\tan\beta$ all the cosmological interesting region will be covered (see text for details).*

The a_μ curve refers to the expected region that will be probed before 2006 by the measurements of the muon magnetic dipole moment ¹⁶⁾. The curve $B \rightarrow X_s \gamma$ refers to the improvement expected for the same date from BaBar, BELLE and B factories in respect to the CLEO and ALEPH results ¹⁷⁾. The curve Φ_μ^\oplus refers to the indirect DM search with underwater ν experiments like AMANDA, NESTOR and ANTARES ¹⁸⁾ and the curve σ_p refers to the direct DM search with underground experiments like DAMA, CDMS, CRESST and GENIUS ¹⁹⁾

We conclude with one last remark, the angular resolution and energy resolution achievable in gamma ray astrophysics is still lower to what is desirable and achievable in other band; so a long term plan like the one sketch in figure 16

Ultimate Objective: To image the particle accelerator near the event

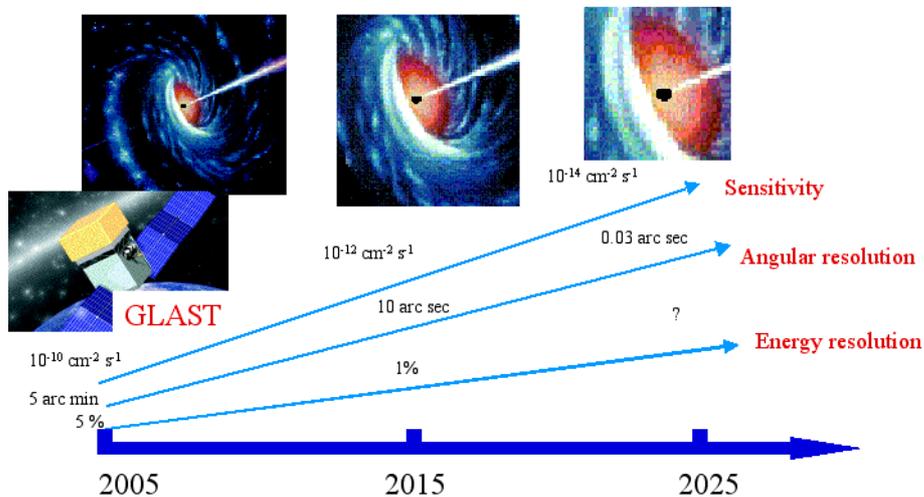


Figure 16: *Gamma-Ray Astronomy Long Term Plan*

is needed and can bring spectacular results.

3 Acknowledgments

I wish to thank all the participants to the school. Everybody (both professors and students) contributes so much to discussions and debates to make the time spent in L'Aquila really very interesting.

References

1. W. Atwood *et al.*, 1994, NIM, **A342**, 302. Proposal for the Gamma-ray Large Area Space Telescope, SLAC-R-522 (1998). B. Dingus *et al.*, 25th ICRC, OG 10.2.17, **5**, p.69, Durban. A. Morselli, Very High Energy Phenomena in the Universe, Ed. Frontiers (1997) 123. A. Morselli, "Frontier Objects in Astrophysics and Particle Physics", SIF, Bologna, **65** (1999) 613.
2. R. Bellazzini, these proceedings.
3. A. Celotti, these proceedings.

4. R.Romani, 1996, *Astrophys. J.* **470**, 469.
5. J.Daugherty and A.Harding, 1996, *Astrophys. J.* **458**, 278.
6. L. Bergström *et al.*, 1998, *Astropart. Phys.* **9**, 137.
7. J. Navarro *et al.*, 1996, *Astrophys. J.* **462**, 563.
8. P.Spillantini *et al.*, 24th ICRC Roma, OG 10.3.7 (1995) 591. V.Bonvicini *et al.*, NIM, **A 461** (2001) 262. P.Spillantini, these proceedings.
9. M. Boezio *et al.*, 2000, *ApJ*, **532**, 653 and references therein.
10. E.Baltz and J. Edsjö 1999 *Phys. Rev.* **D 59**, 023511.
11. P. Picozza and A. Morselli, The Ninth Marcel Grossmann Meeting, World Scientific (2001) [astro-ph/0103117] and references therein.
12. D. Bergström *et al.*, 2000 *ApJ Letters*, **534**, L177 and references therein.
13. P. Ullio 1999, astro-ph/9904086 and these proceedings.
14. P.Fleury , these proceedings.
15. L. Feng *et al.*, *Phys. Rev.* **D 63**, 045024 (2001) [astro-ph/0008115].
16. M. Grosse Perdekamp *et al.*, *Nucl. Phys. Proc. Suppl.* **76** (1999) 253.
17. A. L. Kagan and M. Neubert, *Eur. Phys. J.* **C7** (1999) 5 [hep-ph/9805303].
18. AMANDA Collaboration, E. Dalberg *et al.*, HE.5.3.06, 26th ICRC (1999). C. Spiering *et al.*, astro-ph/9906205. NESTOR Collaboration, <http://www.uoa.gr/nestor>. ANTARES Collaboration, J. R. Hubbard *et al.*, HE.6.3.03 26th ICRC (1999).
19. DAMA Collaboration, R. Bernabei *et al.*, *Phys. Lett.* **B480** (2000) 23. CDMS Collaboration, R. W. Schnee *et al.*, *Phys. Rept.* **307** (1998) 283 . CRESST Collaboration, M. Bravin *et al.*, *Astropart. Phys.* **12** (1999) 107 [hep-ex/9904005].

**AMS- γ : HIGH ENERGY PHOTON DETECTION WITH THE
ALPHA MAGNETIC SPECTROMETER ON THE ISS**

R. Battiston, B. Bertucci, G. Lamanna, B. Alpat, W.J. Burger,
G. Esposito, E. Fiandrini, P. Zuccon
*Dipartimento di Fisica dell' Università and Sezione INFN
Via Pascoli, 06100 Perugia, Italy*

ABSTRACT

AMS is a large acceptance, superconducting magnetic spectrometer designed to study, with high accuracy, the composition of cosmic rays. We discuss how AMS will also identify γ -rays in the energy interval 1 – 300 GeV, operating as a e^+e^- pair spectrometer. During the scheduled three years mission on the International Space Station (ISS) starting in 2004, AMS will provide access to the largely unexplored γ energy range above 20 GeV complementary to other space missions and ground based Cerenkov detectors.

1 Introduction

The successful operation of the Compton Gamma Ray Observatory (CGRO) ⁸⁾ has provided a vast amount of information on the gamma emission in the Universe and has led to the discovery of new phenomena. In particular, the Energetic Gamma Ray Experiment Telescope (EGRET) ⁷⁾ sensitive to high energy

photons (0.1 up to 30 GeV) has been crucial for the identification of several classes of new gamma sources, such as extragalactic blazars, galactic pulsars, γ -ray bursters, as well as providing a measure of the diffuse emission both of galactic and extragalactic origins. A decade of operation has allowed to verify the extreme time variability of such objects and to correlate their high energy photon emission with their counterparts at longer wavelengths. The surprisingly dynamic universe emerging from these measurements has generated much interest in high energy γ -ray detection: new γ -ray telescopes AGILE ⁹⁾ and GLAST ⁶⁾ are currently under construction to extend the survey up to 50 GeV and $\sim 300 GeV$, respectively. In this context, it is interesting to analyze the

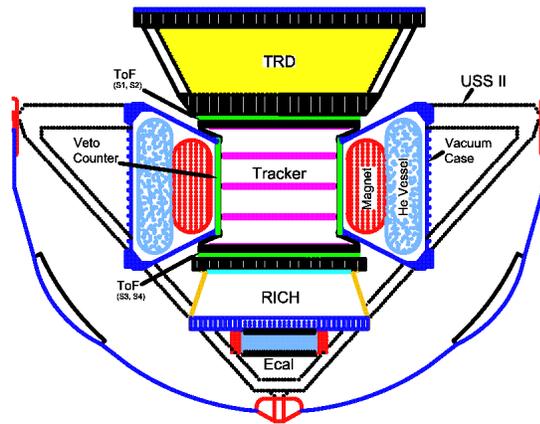


Figure 1: *Schematic view of the AMS-02 experiment which will operate on the International Space Station from the year 2004.*

sensitivity for energetic photon detection of the Alpha Magnetic Spectrometer, a large acceptance particle detector which will be operating on the International Space Station for three years, starting in 2004. The AMS experiment has been proposed to study with high accuracy the composition of cosmic rays. A reduced version of the detector was flown and operated successfully during a 10-day shuttle flight in June 1998. The data accumulated during the flight has provided new accurate measurements of the proton ¹⁾, helium ²⁾ and lepton ³⁾

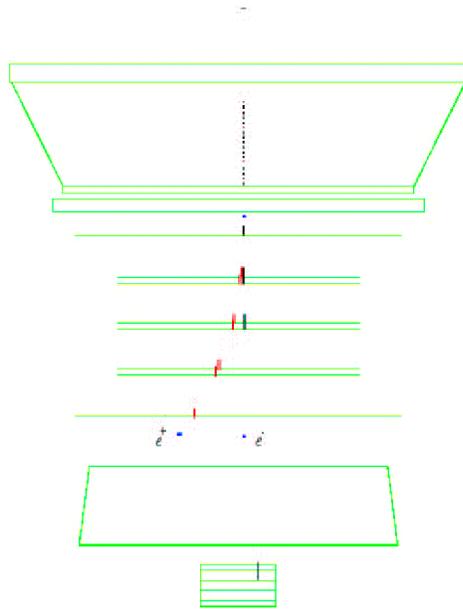


Figure 2: *AMS-02- γ Monte Carlo event: a 10 GeV γ converts in the TRD. The e^- is detected also by ECAL. Such an event is triggered by a scintillator paddle in the first ToF layer and two separated hits on the fourth ToF layer.*

fluxes.

2 The AMS-02 detector

Following the Shuttle flight data analysis, a significant upgrade of the AMS instrument has been started: a substantial increase of the magnetic field strength, the addition of a Transition Radiation Detector (TRD), a Ring Imaging Cerenkov detector (RICH) and an Electromagnetic CALorimeter (ECAL). The present paper represents an update of a precedent study ⁴⁾ of the photon detection potential based on the shuttle-flight version of the detector (AMS-01).

A schematic view of the AMS instrument (AMS-02), in its configuration on the ISS, is shown in Fig.1. It consists of a cylindrical superconducting magnet spectrometer with 1 m inner diameter providing a bending power

$BL^2 \simeq 1 \text{ Tm}^2$, which will allow the detection of nuclei up to rigidities of the order of 10 TV . The track reconstruction will be performed by 8 planes of silicon detectors with a spatial resolution of $10 \mu\text{m}$ ($30 \mu\text{m}$) in the bending (non-bending) projection¹⁰). The inner surface of the magnet will be covered by scintillator counters to veto stray trajectories and background particles. On the two end caps of the magnet two scintillators layers, segmented orthogonally in paddles, will measure the Time-of-Flight (ToF) and provide the fast trigger for the experiment. The TRD, on top of the spectrometer, is 0.6 m deep. Its layout consists of multi-arrays of a 12 mm thin foam radiator seen by two layers of fourteen 5 mm diameter gas proportional tubes. These arrays will provide a measurement of the x and y coordinates of the transition point of the radiated quanta. The TRD will be sensitive to isotropic incident e^- , p and γ , and within the full magnet acceptance and will allow a p/e discrimination up to 300 GeV . The RICH detector will allow light nuclei isotopes separation. Placed immediately downstream the last two ToF planes, it consists of two radiator materials: sodium fluoride (NaF) and aerogel (AGL). A pixel phototubes matrix is used for the light detection. At the bottom of AMS, the ECAL is a three-dimensional electromagnetic sampling calorimeter consisting of 1 mm diameter blue scintillating fibers sandwiched between grooved lead plates. With an overall dimensions of $63 \times 63 \times 16 \text{ cm}^3$ and a total radiation length of $16 X_0$, the ECAL will provide an energy resolution $\sigma/E \sim (5.5/\sqrt{E(\text{GeV})} + 2.5)\%$ providing a proton suppression factor of about 10^{-4} up to 100 GeV .

3 The simulation

The response of the AMS for γ -ray detection by means of the identification and reconstruction of e^+e^- pairs from conversions, in the material upstream of the second silicon tracker layer, was studied with the AMS detector simulation and reconstruction program based on GEANT⁵). Gamma rays have been generated isotropically over the full solid angle of the detector, at several fixed energies, ranging from 1 to 300 GeV . All physical processes for electrons and γ -rays were included. However the bremsstrahlung photons were not followed for energies below 20 MeV . The material in front of the first silicon tracker plane, consisting of the TRD, the first two layers of ToF scintillators and mechanical supports, represents $\simeq 0.22 X_0$.

Events with photons converting in the detector material and releasing a

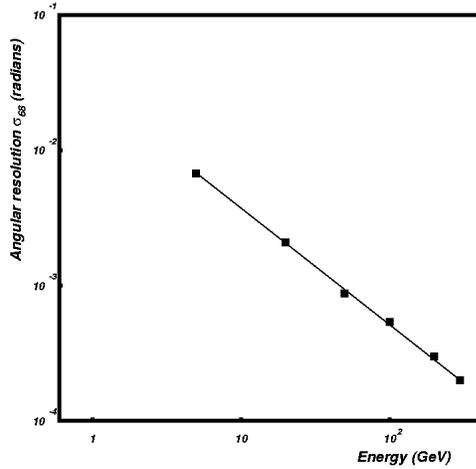


Figure 3: *Angular resolution of AMS-02- γ as a function of primary energy in the interval 5 to 300 Gev.*

signal in at least two scintillator planes were considered for the reconstruction of the e^+e^- pair in the spectrometer. A minimum of 4 reconstructed hits over a minimum lever arm larger than 3/4 height of the tracker (from the first to the last plane) for each lepton track was required in order to accept the event. Primary γ -ray energy and incidence direction were determined by adding the fitted momenta vectors of all secondaries, evaluated at the entrance of the magnetic field. An example of a reconstructed MC event is shown in Fig.2.

4 The Results

4.1 Sensitivity

Astrophysical γ -sources fall into two categories: point sources, such as blazars and GRBs, and diffuse sources, such as the cosmic isotropic γ -ray background, as well as possible γ -ray emission from neutralino annihilation near the center of our Galaxy. Some galactic sources fall into an intermediate category of "extended" sources, such as pulsar nebula and supernova remnants. The capability to disentangle a source signal from background depends on several factors: the angular and energy resolutions, the detector aperture and the exposure time at

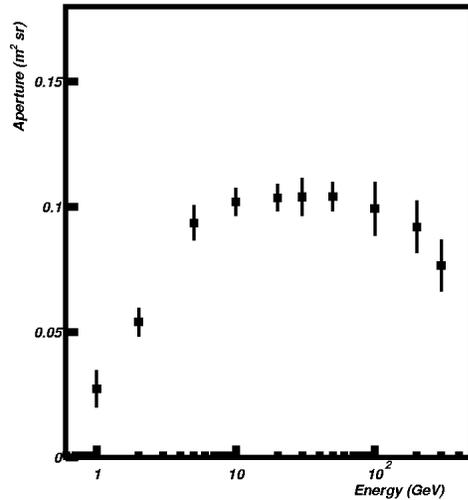


Figure 4: *AMS aperture* $A(E, \theta)$ *as a function of* γ -*ray energy. The errors are statistical.*

a given position on the celestial sphere. Following EGRET's definition ¹¹⁾, the solid angle $\Omega(E)$, over which the background must be integrated when viewing a source, is $\Omega(E) = \pi\sigma_{68}^2(E)$, where $\sigma_{68}^2(E)$ is defined as the angular radius within which 68% of the source photons fall ⁴⁾. The results of AMS-02- γ simulation with the EGRET criteria yield the angular resolutions as a function of photon energy shown in Fig.3. They range from 10^{-4} to 10^{-2} *rad*, due mainly to the multiple Coulomb scattering.

The second figure of merit for photon detection is the detector aperture. The AMS-02- γ aperture has been determined as a function of the γ -ray energy between 1 and 300 *GeV* (Fig.4). The aperture is a result of three main contributions: the AMS geometrical acceptance ($0.5 \text{ m}^2 \text{ sr}$), the pair conversion probability, and the double-track reconstruction efficiency. As seen in Fig.4, below a γ -ray energy of 5 *GeV*, the converted electrons have a too small radius of curvature for detection. Above 200 *GeV* the electron and positron detection is limited by the double-hit resolution of the silicon tracker. The AMS-02- γ aperture is also a function of the inclination of the incident photon, dropping to zero at about 50° as it is shown in Fig. 5. A second selection was also applied to estimate the acceptance for photons detected only in the electromagnetic

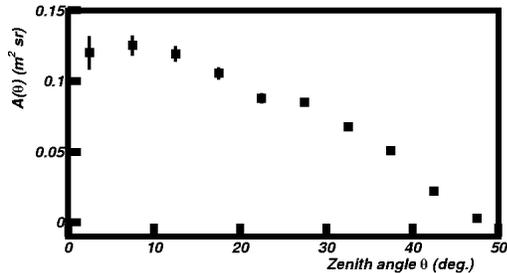


Figure 5: *Differential aperture versus zenith angle at 50 GeV.*

calorimeter. We found that this approach would increase the AMS-02- γ acceptance by a factor of 2, attaining 0.2 to 0.3 $m^2 sr$ as a function of the energy, with a maximum energy between 100 and 200 GeV . There is a possibility that the ECAL signal can be implemented in the trigger of the experiment.

4.2 Energy Resolution

The γ -ray energy resolution is dominated by bremsstrahlung losses throughout the material traversed by e^+e^- pairs after the conversion. The fractional energy loss of the leptons is proportional to x/X_0 . As a consequence the energy resolution depends on the distance between the conversion point and the position of the second silicon tracker plane.

Figure 6 shows the reconstructed γ -ray energy distributions for primary energies of 5, 50, 100 GeV . One sees the effect of bremsstrahlung losses in the large low-energy tails of the distributions. The nearly identical appearances of these distributions is a consequence of the fact that the probability for the electron to radiate a fraction ν of its energy as a bremsstrahlung photon depends only on ν . There are two additional effects in the distributions in Fig. 6: the energy cutoff at 1 GeV due to the minimum detection threshold of the momenta of the leptons, and the broadening of the distribution at 100 GeV due to the fact that trajectory reconstruction errors begin to dominate bremsstrahlung

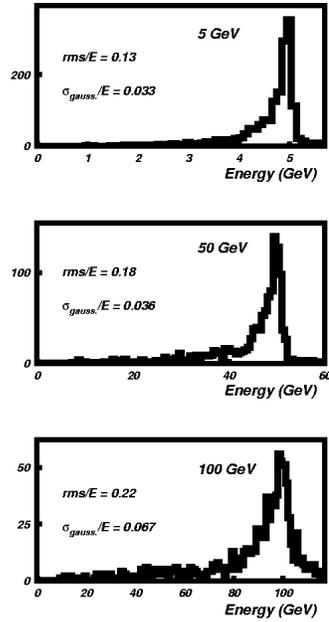


Figure 6: The reconstructed energy distributions for three different primary γ -ray energies, 5, 50, 100 GeV. The r.m.s. energy resolution and the width of the "gaussian" peak of the distributions are also indicated.

losses in the energy determination. In Fig. 6 the r.m.s. energy resolution and the estimated "gaussian" width of the "peak" are also indicated for the three different primary γ -ray energy.

5 Conclusions

The performance of AMS-02 for γ -ray detection, via pair production, has been studied. Our preliminary results confirm the previous study ⁴⁾ based on AMS-01 design, i.e. the performance in terms of acceptance, angular and energy resolutions makes AMS-02 a good candidate for gamma-ray detection. Furthermore the decision to incorporate an electromagnetic calorimeter in AMS-02 enhances the gamma potential detection of the device.

This work has been partially supported by the Italian Space Agency (ASI) under contract ARS-98/47.

References

1. AMS Collaboration, Alcaraz, J. et. al., *Phys. Lett.* **B472**, p. 215, 2000.
2. AMS Collaboration, Alcaraz, J. et. al., *Phys. Lett.* **B484**, p. 10, 2000.
3. AMS Collaboration, Alcaraz, J. et. al., *Phys. Lett.* **B494**, p. 193, 2000.
4. Battiston R. et al., *Astrop. Phys.* **13**, p.51, 2000.
5. Brun R. et al., *GEANT3*, CERN-DD/EE/84-1 (Revised 1987).
6. Engovatov D. et al., *IEEE Nucl. Sci. Symp. (Albuquerque)*, 1997.
7. Hartman, R.C. et al., *Astro. Jour. S.* **123**, p.79, 1999.
8. Kurfess J.D. et al., *AIP Conf. Proc.* **410**, p.509, 1997.
9. Morselli A. et al., *Nucl. Phys. B (Proc. Suppl.)* **85**, p.22, 2000.
10. Rapin D., *Proc. ICRC* , 2001.
11. Thompson D.J. et al., *Astro. Phys. Journ. Supp.* **86**, p.629, 1993.

The Physics of Active Galactic Nuclei

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

ABSTRACT

A brief introduction on what (we hope) believe we might understand of the physics of Active Galactic Nuclei (AGN) will be presented. Starting with a broad definition of AGN, the hypothesis on the functioning of the central engine, the main processes responsible of the observed radiation properties, we then concentrate, because of the limitation of space and the main focus of this School, on radio-loud AGN which emit an energetically significant, and even dominant, radiative contribution in the γ -ray band.

1 What are they?

AGN are characterized by and classified according to a wealth of observational properties, but broadly speaking they are identified as non stellar excess emission from the very center of at least 10 % of galaxies ¹⁾, ²⁾, ³⁾.

In a large fraction of AGN the high emitted luminosity L (say between $10^{42} - 10^{48}$ erg s^{-1}) and the small typical size R (as inferred from variability timescales $R < t_{var}c$) imply that the conversion of mass into radiation has to be highly efficient, and cannot be due to nuclear fusion. This led to postulate that on such astrophysical scales gravity could release more energy: if the gravitational potential is that of black holes of $\sim 10^6 - 10^9 M_{\odot}$ the conversion of gravitational energy of matter in such a deep potential could give origin to the AGN power. More precisely the theoretical maximum extractable energy corresponds to an efficiency up to 42% for a spinning black hole (see e.g. 1), 4), 5)). Indeed masses in the range $\sim 10^6 - 10^9$ solar masses can be inferred from a simple and general estimate by requiring that, if accretion has to be effective, the associated Eddington luminosity exceeds the luminosity typically observed. Furthermore these same two quantities, L and R , also imply large values of the compactness, $\sim L/R$, in the central regions, suggesting a nuclear environment where processes involving interaction with photons are highly effective.

However an important issue follows: as matter/gas on galactic scales is expected to have an angular momentum largely exceeding that required to reach orbits close to the hole, most of the angular momentum has to be lost for the mass feeding to proceed. On scales already within the influence of the black hole potential, the angular momentum dissipation is most commonly assumed to occur via viscosity between gas differentially rotating in keplerian orbits in a (thin) disc like configuration. The energy is supposed to be then released as quasi-thermal radiation by the heated optically thick gas. Other configurations and physical conditions of the accreting matter (such as quasi-spherical structures of low density and low angular momentum gas) have been proposed and discussed (see e.g. 6)).

While several pieces of evidence accumulated in favour of the presence of supermassive black holes at the center of Active Galaxies 1), 5) (residual mass from accretion, presence of straight long lived jets, dispersion velocity of stars in the core of the host galaxies, width of iron lines in the AGN X-ray spectra, etc), only recently this has been (possibly undoubtedly) indicated by measures of stars and gas dynamics at high spatial resolution: these consistently show the presence of a supermassive object at the center of most (if not all) galaxies and this has to be so small that a black hole becomes the most plausible possibility. Intriguingly, these studies have then found a direct and quite tight link

(proportionality) between the mass of the hole and properties connected with the bulge of the host galaxy, e.g. 7), 8), a fundamental piece of information to understand the formation of the black hole/galaxy system.

In such a scenario, of an ubiquitous presence of black holes at the center of galaxies, the AGN phenomenon then corresponds to a phase where (enough) gas has reached the vicinity of the black hole and can be effectively accreted. The role of interaction and merging between galaxies in facilitating the gas to reach the nuclear galactic regions is unclear. Paradoxically, we are even faced with the fact that many galactic nuclei, where we can detect the presence of gas, are less active than simply predicted by our theories.

Furthermore, it is worth stressing that on top of the AGN non-thermal emission, it has become increasingly evident that the nuclear regions can be also associated to the presence of copious star formation, which is at least in part responsible for unusual environments and conditions in the galactic nuclei.

2 What do they look like?

Let us now consider more specifically what we observe from AGN and how this can be accounted for in the general framework just described. To avoid a long and detailed description of the AGN zoology, let me only consider one main characteristic which distinguish two broad classes, the so called radio quiet (RQ) and radio loud (RL) AGN. These are distinguished on the basis of their radio vs optical flux ratio, where in the latter ones it (clearly) assumes larger values (by about one-two orders of magnitude). Although the RQ/RL might not be a clear dichotomy as once thought, this division can help understanding the nature of the sources, as the radio loudness is associated with the presence of collimated jets of plasma, symmetrically ejected from the nuclear engine and extending in the two opposite directions up to scales of hundreds kpc 9), 10).

In general, the broad band spectral energy distribution (SED) of AGN can extend from radio to γ -ray energies, and it is typically non-thermal (i.e. not a singly peaked spectral distribution). More specifically, the SED of RQ sources are clearly characterized by three broad bumps (in a $\nu L(\nu)$ representation), namely in the IR, optical-UV and hard X-ray bands. The RL population might be described as two smooth very broad peaks superposed to the RQ SED, which dilute or even completely mask the three RQ bumps: the lower energy peak connects to the (excess) radio emission, and the higher energy one

may extend up to GeV and TeV energies.

Finally, at least another important observational information should be mentioned, namely the presence of emission lines in the IR-UV spectra of most AGN. In some objects both broad (FWHM of several thousands of km s^{-1}) and narrow (few hundreds of km s^{-1}) lines are detected, while in other sources only the narrow lines are present. These lines are believed to originate in gas clouds located at different distances from the photoionizing nuclear source, their width being associated to the (quasi-keplerian) dynamical motion.

3 How do they emit?

Let us now consider the origin of the above spectral components and what can be inferred on the structure and physical conditions of the nuclei.

The main component evident in the spectrum of RQ sources as a bump (excess) of optical-UV photons is believed to be the direct signature of quasi-thermal radiation from the accreting gas. Indeed its temperature of few tens of thousand degrees most naturally corresponds to the typical temperature of gas in thermal equilibrium located in the nuclear regions and emitting the observed luminosity ($L \sim 10^{45} - 10^{46} \text{ erg s}^{-1}$) like a (quasi-)blackbody.

The IR bump is instead thought most likely to originate on larger scale dusty structures. These would intercept part of the UV photons and re-emit in the IR band. This hypothesis is strongly supported by the commonly observed minimum in the SED at $\sim 1\mu\text{m}$, corresponding to the temperature of sublimation of dust.

Finally let us briefly consider the X-ray component. The best studied X-ray spectra are of course those of low redshift bright sources (typically Seyfert galaxies). The analysis of such spectra reveals that the spectrum can be well interpreted as due to the contribution of two fundamental components. A 'primary' spectrum, well described by a power law photon distribution, likely originating as Comptonization of soft (disk?) photons by thermal plasma, at few hundred keV, forming a corona above the disk. The second one consists of a component peaked at few tens keV, which plausibly arises from the impinging of the primary spectrum onto an optically thick surface (most likely the accretion disk structure): the 'reflected' spectrum, which reaches the observer, would in fact resemble a broad peak at few tens keV (lower energy photons are photoelectrically absorbed and higher energy ones are downscattered in the

optically thick medium) plus the emission features from ionized atoms (most prominent are iron features, i.e. lines and edges below ~ 10 keV).

4 Radio-loud sources

As mentioned, radio loud AGN show, superposed to these components associated more or less directly with the nuclear accretion phenomenon, the contribution of non-thermal emission characterized by a smooth continuum. This highly variable radiation is believed to be emitted by the plasma flowing in the jets which connect the nuclei to the extended (~ 100 kpc) structures. This radiation is observable over the broadest energy range, at least partly covering the dust and optical-UV bumps and the X-ray continuum, and completely dominating in the radio and - when visible - in the γ -ray bands. In fact, the EGRET instrument onboard the Compton Gamma-Ray Observatory detected several tens of sources copiously emitting in the \sim GeV range, which have been soon identified with radio-loud AGN. A few sources (about four to date) have been also seen emitting in the TeV range by Cherenkov telescopes.

An important physical property of the plasma in the (inner parts of) jets is that its bulk motion is at highly relativistic speeds. By now there is compelling evidence for relativistic motion, which is a crucial fact to take into account when interpreting several and independent observed phenomena. These include: a) measured brightness temperatures of radio sources exceeding the limit for catastrophic inverse Compton cooling; b) the observation of radio components on the pc scales apparently moving at speeds exceeding the speed of light; c) the detection of highly energetic γ -rays which can thus be able to avoid absorption in the source itself (via photon-photon interaction leading to pair production); d) the one sidedness of many jets, etc.

How can these observations be accounted for by relativistic motion? The relativistic motion implies that an observer located at a small angle with respect to the velocity direction would observe (because of photon aberration, plus energy and time transformations) a highly enhanced luminosity, much larger than the isotropically emitted comoving one. By considering this effect it is then possible to lower the expected comoving radiation density and thus reduce the above (and more) difficulties [the simplest explanation for superluminal motion actually only considers the 'ordinary' Doppler effect, dominating at small angles and high velocities].

Sources dominated by emission from the highly relativistic moving plasma are called blazars (including both BL Lac objects and flat spectrum radio quasars) and, as already mentioned, their spectra can be well described by two broad components (in a $\nu L(\nu)$ representation). The first component, peaking between the IR and X-ray band, is produced by synchrotron emission of a non-thermal distribution of relativistic electrons, as indicated by the non-thermal character of the spectrum and the high radio and optical polarization. The origin of the second component, which often energetically dominates the radiative output, is instead not clearly and univocally established. One possibility is that it arises as inverse Compton emission - by the same non-thermal electron distribution - on the synchrotron photons themselves or on other photons produced outside the jet (such as disk emission, broad line photons, etc.): the latter 'external' photon fields in fact would be seen enhanced by the emitting plasma due to its relativistic motion. A second class of 'hadronic' models instead invoke synchrotron emission from either relativistic protons or particles resulting from the interaction of the protons with either lower energy protons or photons. As the former ('leptonic') class of models appears to be more widely accepted, I will hereafter focus on its consequences on the physics of blazars.

In fact, the construction and modeling of the energy distribution of a large number of blazars have revealed an interesting trend among the frequency position of the synchrotron peak, the relative ratio of the high energy (inverse Compton) and lower energy (synchrotron) components and finally the total radiated power ¹¹⁾. This trend can be tantalizing ascribed to a systematic increase in the soft external photon field with power which leads to a relatively more prominent inverse Compton emission and lower typical energy of the electrons cooling at the peak of the distribution ¹²⁾. This simply accounts for global properties of the whole of the blazar family from the lower power BL Lac (which do not require an external photon field) to the more powerful flat spectrum radio quasars (where the external field dominates).

As clearly emerges from the above, the study of γ -ray emission from blazars is crucial, as it often corresponds to the bulk of the radiative dissipation of the jet, being thus crucial in estimating its effective power. GLAST will be able to increase both the number of objects detected and the quality of their γ -ray spectra. As the emission in these sources is highly variable also an instrument like AGILE could be extremely effective, in particular in estimating

the duty cycle of γ -ray activity in a significant number of sources. This in fact is directly linked to the physics of the jet formation and propagation and has also important consequences on the estimate of the contribution of blazars to the γ -ray background. Further important (cosmological) results are also expected from the detection in a significant number of blazars of absorption of TeV photons by the soft photon bath of the infrared background with the new generation of Cherenkov telescopes.

5 The unifying view

Unification models try to disentangle effects which only depend on geometry and orientation from intrinsic and physical ones.

In this respect two aspects have been considered so far. The first one concerns the observation of sources which only show the presence of narrow lines in their optical-UV spectra. In fact spectropolarimetry of some of these sources (mainly of Seyfert galaxies) has revealed that also broad lines are in fact visible in polarized light¹³. These objects, named as type 2 (to be distinguished from the broad lined type 1) also typically show absorption column densities in excess of the Galactic value in the soft X-ray band. These two properties can be accounted for at the first order by the existence of absorbing structures of dust and gas which in some objects obscure our line of sight to the central nucleus (corresponding to the type 2 sources) while in other cases the nucleus can be seen directly (the case of type 1 objects). The structure, extending on scales larger than pc, is usually referred to as a torus (though there are no indications on its actual geometrical distribution). In this way, type 1 and type 2 objects would be unified into one class of intrinsically identical sources.

The second anisotropy which indeed requires unification is that associated with the relativistically beamed emission from blazars. Once again the models try to identify the so called 'parent population', i.e. the class(es) of known AGN which are blazars observed at larger angles with respect to the jet plasma motion and thus not strongly affected by the relativistic beaming of the radiation. According to the most accredited unification scenarios low (FR I) and high (FR II) power radio galaxies might constitute the parent populations of low (BL Lac) and high (flat spectrum radio quasars) power blazars¹⁴.

Having excluded the geometrical effects due to orientation, the essential physical parameters determining the fundamental properties of AGN might

reduce to only three, namely the black hole mass, the rate at which gas is accreted onto the hole releasing part of its gravitational energy, and possibly the spin of the hole. The latter one has been in fact invoked as plausible source of power for jets, thus possibly accounting for the different radio loud/radio quiet phenomenologies, but the issue is still far from being clear ⁹, ¹⁰).

6 Acknowledgments

I thank the Organizers for having invited me to such a stimulating School. It is a pleasure to thank Gabriele Ghisellini for a critical reading and years of fruitful collaboration. The ISSS and the Italian MUIR are acknowledged for financial support. I apologize with colleagues for the paucity of references given, due to the available space for such a vast subject.

References

1. R.D. Blandford *et al* SAAS-FEE Course 20 (Springer-Verlag, 1990)
2. I. Robson, Active galactic nuclei (Wiley-Praxis Publishing, 1996)
3. B.M. Peterson, An introduction to active galactic nuclei (CUP, 1997)
4. M.J. Rees, ARA&A **22**, 471 (1984)
5. J.H. Krolik, Active galactic nuclei (Princeton University Press, 1999)
6. J. Frank, *et al*, Accretion power in astrophysics (CUP, 1992)
7. J. Kormendy, D. Richstone, ARA&A **33**, 581 (1995)
8. L. Ferrarese, D. Merritt, ApJ **539**, L9 (2000)
9. M.C. Begelman, *et al*, Rev. Mod. Phys. **56**, 255 (1984)
10. Astrophysical jets, eds. D. Burgarella *et al* (CUP, 1993)
11. G. Fossati *et al*, MNRAS **299**, 433 (1998)
12. G. Ghisellini *et al*, MNRAS **301**, 451 (1998)
13. R. Antonucci, ARA&A **31**, 473 (1993)
14. C.M. Urry, P. Padovani, PASP **107**, 803 (1995)

THE SIMULATION OF ASTROPARTICLE EXPERIMENTS AND THE GEANT4 TOOLKIT

Alessandro De Angelis

Dipartimento di Fisica dell'Università di Udine and INFN Trieste

ABSTRACT

The new generation of astroparticles experiments will use technologies common to High Energy Physics. Among such technologies a central place is given to the role of simulation.

1 Introduction

Monte Carlo simulation is nowadays an essential tool in the project of experiments. It is so common for physicists that one could think that such a technique is in use since ages; but indeed it is quite young, around 50 years. It was suggested by Ulam and von Neumann ¹⁾ in 1947 and first used for particle transport by Wilson ²⁾, in the context of the the problem of electromagnetic showers. Such a problem had an elegant solution by Rossi and Greisen ³⁾, although based on approximations valid beyond energies of a few MeV.

To solve the integro-differential equations of showers in lead, Wilson in 1952 uses the following procedure.

“The procedure used was a simple graphical and mechanical one. The distance into the lead was broken into intervals of one-fifth of a radiation length (about 1 mm). The electrons or photons were followed through successive intervals and their fate in passing through a given interval was decided by spinning a wheel of chance; the fate being read from one of a family of curves drawn on a cylinder (...) A word about the wheel of chance. The cylinder, 4 in. outside diameter by 12 in. long is driven by a high speed motor geared down by a ratio 20 to 1. The motor armature is heavier than the cylinder and determines where the cylinder stops. The motor was observed to stop at random and, in so far as the cylinder is concerned, its randomness is multiplied by the gear ratio (...)”

What are the requirements of a general-purpose simulation software for particle transport? To start, one has to take care in an adequate way of the physics, i.e., of the interactions: both for what is related to the probability of occurrence (i.e., to the cross section) and to the modelling of the final state. The physics of electromagnetic interactions is based on QED, and thus it is in principle well known. Electromagnetic interactions should be well modelled down to low energies (how low depends on the geometry and on the physics of the detector). The theory of hadronic interactions, QCD, is not in a status comparable with QED; as a consequence the simulations of hadronic interactions relies in general on QCD-inspired models rather than on the theory itself, and in general a reasonable simulation of hadronic interactions is good enough.

For a software to be used in the context of detector project and data analysis, however, the physics requirements are not the full story, and maybe they are not even the most important part. Technical requirements rely mostly on a well written code, with characteristics of

- modularity;
- easiness to add different generators;
- easiness to add new physics routines;
- friendly interfaces;
- good documentation;
- maintainability;

- support on different platforms;
- last but not least, a term taken from industry: customer care.

2 The shop-list before Geant4

Most of the large experiments '90s were basing their simulation on the Geant3 ⁴⁾ package. The Geant3 code comes from a long program of development based at CERN between 1982 and 1994. Geant3 is written in FORTRAN; among his strong points, compared to the other simulations, are user-friendly packages for geometry description and visualization, and an overall easiness-of-use. Geant3 has proprietary routines for electromagnetic physics, and can bind several hadronic codes (GHEISHA is the most common).

However, a different package was and still is the reference for electromagnetic physics: EGS ⁵⁾, a package which had a very long development and debugging at SLAC, LNL and KEK in the period between 1966 and 1985. EGS, presently at the version 4 (EGS4), is still the reference now for dosimetry, where definitely one should not be wrong. EGS4 has proprietary routines for electromagnetic physics; it is most commonly used in couple with FLUKA for hadronic interactions. EGS4 is written in MORTRAN, a pre-processor of FORTRAN, although several FORTRAN and C++ versions circulate on the Web. The code is overall a bit unfriendly: the geometry is difficult to define (it must be put in by means of a subroutine); few facilities, in particular for visualization, are available. The cross sections are computed by an offline preprocessor, PEGS, which must be run separately.

Besides Geant3 and EGS4, very little room was available for other products (DELSIM, GISMO etc.); and the success of a simulation code is boosted by the widespread use (which guarantees updating, debugging, availability on different platforms, interfaces).

3 The Geant4 toolkit

At the end of the '90s most of the simulation programs in High Energy Physics were based on Geant3. Several reasons however pushed the community to start a new project, aimed to improve Geant3. Among such reasons:

- limitations of Geant3 maintenance; in particular, because of too complex

structure driven by historical reasons, it was almost impossible to add a new feature or to hunt a bug;

- limitation of FORTRAN, and choice of object orientation and C++ by the physics community;
- shortage of man power at CERN;
- limitation of “central center” supports.

Such reasons convinced a group of physicists and computer scientists at CERN to start a world-wide collaboration for a new simulation project. Such a project was based on the most recent software engineering methodologies and, from an organizational point of view, on a world-wide collaboration.

3.1 Geant4: philosophy, history, future

Geant4 aims to be the successor of Geant3 by redesigning a major package of CERN software for the next generation of HEP experiments using an Object Oriented philosophy. The final aim of the project is to build a simulation more precise than EGS, and more friendly than Geant3.

A variety of new requirements also came from heavy ion physics, CP violation physics, cosmic ray physics, medical applications and space science applications. In order to meet such requirements, a large degree of functionality and flexibility are provided: Geant4 is not only for High Energy Physics.

The main steps of the history of Geant4 are, for the time being:

Dec 94 Project starts

Apr 97 First alpha release

Jul 98 First beta release

Dec 98 Release 0.0

Jun 01 Release 3.2 (with a complete list of physics processes)

Maintenance and upgrade are expected for at least 10 years; development is continuous, with two major releases each year plus a monthly internal tag (frequent bug fixes, new features, new examples).

Geant4 is presently based on more than 700,000 lines of code.

For more details on the collaboration and an updated status, see the very well done and maintained web pages ⁶⁾.

3.2 The physics of Geant4

The library proposes several models for the most important interactions of particles with matter.

In particular for electromagnetic interactions one can use a “standard” package with at least all the features of Geant3, or call a package specialized in the low energy part (which aims to an accurate modelling of physics down to 250 eV and below, based on an important use of experimental tables). On the other direction Geant4 should be able to reproduce interactions up to the PeV and beyond: Geant4 is developed by people involved in fields other than High Energy Physics (e.g. medical physics, astroparticle physics).

All processes are already at least at level of Geant3, and in addition there are new processes (transition radiation, optical processes) and substantial improvements have been done in particular on the multiple scattering (new model, without path length restriction and with lateral displacement), on energy loss and on hard processes: in the future also the photoproduction of hadronic resonances will be modeled.

The electromagnetic processes are going through extensive tests, comparing both with data and with Geant3-based and EGS4-based simulations. Very good agreement with the data is seen on the simulation of electromagnetic showers.

For what is related to hadronic interactions one can bind GHEISHA, but more performant models have just been released for the Geant4 code and are under test. There is a large variety of models according to the energy, including string models (Geant4 is interfaced with Pythia7 for hard scattering), cascade models, evaporation and break-up.

In any case Geant4 is an open system to new inputs: the framework is such that different models can be easily integrated.

3.2.1 Confrontation with data

Facing such a huge development effort, debugging and tests is a major enterprise.

Fortunately, the distributed organization helps in boosting the manpower. Many comparisons have been done, and results published; a lot more are ongoing or starting within the collaborations using Geant4, in particular the LHC collaborations (ATLAS, CMS, LHCb, ALICE), BaBar (migrating from Geant3) and GLAST.

Again the most important results are summarized in the Geant4 web page.

3.3 Miscellaneous features

Geant4 includes several features going in the direction of functionality and easiness of use; many such features come from comments originated from Geant3 users.

In particular such improvements affect the definition of cutoffs, geometry and utilities, hits and digitizations, fast simulation, visualization.

3.3.1 Cutoffs

Contrary to what done in Geant3, cuts in Geant4 are done in range rather than energy. It makes poor sense to use the energy cut-off: for example the range of a 10 keV gamma in Si is a few cm, while the range of 10 keV electron is a few micron.

This modification causes a significant gain in results quality versus CPU usage; however, users can override the default and impose a cut in energy, track length, or time-of-flight. Physics processes can also ask to override the default when they need to (for example for a better treatment of boundary effects).

3.3.2 Geometry

Like in the philosophy of Geant3, Geant4 pre-defines basic geometries. The user can build new solids from union, intersection, subtraction of two solids (boolean solids) plus a transformation.

A utility *g3tog4* is provided to convert a Geant3 geometry into Geant4.

An interface with XML is in progress.

3.3.3 Hits and digits

Each logical volume can have a pointer to a sensitive detector; a hit is a snapshot of the physical interaction of a track or an accumulation of interactions of tracks in the sensitive detector.

A sensitive detector creates hit(s) using the information given in a tracking step; the user has to provide his/her own implementation of the detector response. A digitization is created with one or more hits and/or other digits by an explicit implementation by the user.

3.3.4 Fast simulation

Geant4 allows to perform full simulation and fast simulation (based on shower parametrizations, less accurate cutoffs etc.) in the same environment. The fast simulation produces the same objects as the full simulation (tracks, clusters etc.)

The full design is such to guarantee flexibility: the user can activate fast/full simulation by detector and/or by particle type, and use parallel geometries.

3.3.5 Visualization

Geant4 provides interfaces to graphics drivers (DAWN, RayTracer, OPACS, OpenGL, OpenInventor, VRML) such that one can visualize detector, hits and trajectories.

3.4 Things one has to do to run Geant4

Documentation (Getting started and installation guide, User guide for application and toolkit developer Software and physics reference manuals) is available at the Geant4 web site.

For many users, however, starting from the study of a manual is not the most effective way. For such users examples are provided: they can go to the Geant4 Web site, run an example and see how it is done.

Six novice examples are available with simple detectors and different experiment types to demonstrate the essential capabilities of Geant4: transport of a non-interacting particle through a slab, track in a simplified tracking detector,

electromagnetic shower (full), particle collision, parametrised electromagnetic shower, optical photon.

In addition advanced examples are available, two of which are relevant for astroparticle:

- `xray_telescope`, illustrating an application for the study of the radiation background in a typical X-ray telescope;
- `gammaray_telescope`, illustrating a gamma satellite-based detector of the new generation, similar to AGILE and GLAST ⁷⁾.

3.5 Experience with Geant4

The production release is in use by many experiments in High Energy Physics and Astroparticle and by groups involved in medical physics.

Thanks to the wide use, the Geant4 developers got feedback. The first results confirm some of the Geant4 strengths in performance, simplicity of use, electromagnetic physics.

Benchmarks between Geant3 and Geant4 in electromagnetic showers demonstrate that Geant4 gives better physics at the same speed (and better speed for same physics).

The tests evidence also some weaknesses; reaction is fast.

4 The simulation of GLAST

An example of implementation for space applications is the simulation of the GLAST gamma-ray telescope ⁸⁾.

GLAST has a wide range of physics objectives, from gamma astrophysics to fundamental physics. Correspondingly, the simulation should have an easy interface to the simulation of different sources, and be adequate both for the design and the physics analysis.

In addition, the gamma simulation in the tracker and in the calorimeter needs different details, and in particular a fast simulation should be available for the huge hadron background.

The GLAST simulation has been done, from the beginning, using C++ and with OO technologies in mind (GISMO was the choice, also because no other candidate present at that moment apart from standard FORTRAN simulations).

Figure 1: *Display of the interaction of a gamma ray with the detector in the Geant4 simulation of GLAST.*

The migration of the GLAST simulation to Geant4 is now almost complete; it uses a prototype of the XML interface for geometry description.

5 Conclusions

Geant4 has demonstrated to be suitable as a Monte Carlo toolkit, in particular for applications in astroparticle physics and High Energy Physics. Among its strong points are the open structure (making it easy to integrate with specialized software) and the easyness of use.

The communities of astroparticle physics and High Energy Physics are quickly acquiring a good experience, and the validation with data, formulae and standard simulations is progressing fast.

In conclusion, Geant4 is becoming the standard *de facto* both for the simulation of detectors and for particle and radiation transport.

Acknowledgements

Many thanks to Riccardo Giannitrapani, Francesco Longo and Maria Grazia Pia for material and useful discussions.

References

1. S. Ulam and J. von Neumann, *Bull. Am. Math. Soc.* **53**, 1120 (1947).
2. Wilson, *Phys. Rev.* **86**, 261 (1952).
3. B. Rossi and K. Greisen, *Rev. Mod. Phys.* **13**, 240 (1941).
4. R. Brun et al., “GEANT Detector Description and Simulation Tool”, CERN Program Library, W5013, 1994.
5. <http://www.slac.stanford.edu/egs/>
6. <http://Geant4.web.cern.ch/Geant4/>
7. A. De Angelis et al., “Simulation for astroparticle experiments and planetary explorations: tools and applications”, to be published in the Proceedings of CHEP 2001, Beijing 2001.
8. <http://glast.gsfc.nasa.gov/>

MEGA - A NEW TELESCOPE FOR MEDIUM ENERGY GAMMA-RAY ASTRONOMY

Gottfried Kanbach

Max-Planck-Institut Extraterrestrische Physik, 85741 Garching, Germany

R. Andritschke, P.F. Bloser, F. Schopper, V. Schönfelder, A. Zoglauer

Max-Planck-Institut Extraterrestrische Physik, 85741 Garching, Germany

ABSTRACT

A new telescope for Medium Energy Gamma-Ray Astronomy (MEGA) is being developed as a successor for the instruments COMPTEL and EGRET on CGRO (1991-2000). MEGA is designed to improve the sensitivity for astronomical sources in the energy range 0.4-50 MeV by at least an order of magnitude with respect to the past instruments. It will fill a severe 'sensitivity' gap between already scheduled hard-X-ray and high-energy γ -ray missions. Key science objectives for MEGA are the investigation of cosmic high-energy accelerators and of nucleosynthesis sites with γ -ray lines. The large-scale structure of the galactic and cosmic diffuse background is another important goal for this mission. MEGA records and images γ -ray events by completely tracking Compton and pair creation interactions in a stack of double sided Si-strip track detectors and 3-D resolving CsI calorimeters.

1 Introduction

After a decade of high-energy astronomy with CGRO (1991-2000), which provided nearly complete spectral coverage extending from several 10 keV to above 30 GeV, the next generation of missions is in preparation. INTEGRAL, to be launched in 2002, will focus on pointed observations of line and continuum sources up to 10 MeV. The Italian AGILE project in ~ 2003 , and the GLAST mission in 2006 will take up astronomy in the energy range above ~ 50 MeV. At medium energies, above several 100 keV and up to ~ 50 MeV, concepts for advanced Compton telescopes, e.g. the Advanced Compton Telescope (ACT) mission with a sensitivity of $\sim 10^{-7}$ photons $\text{cm}^{-2} \text{s}^{-1}$ (a factor 100 improvement over COMPTEL) are discussed in NASA studies; but these ambitious plans lie at least a decade in the future. For the near future, overlapping with

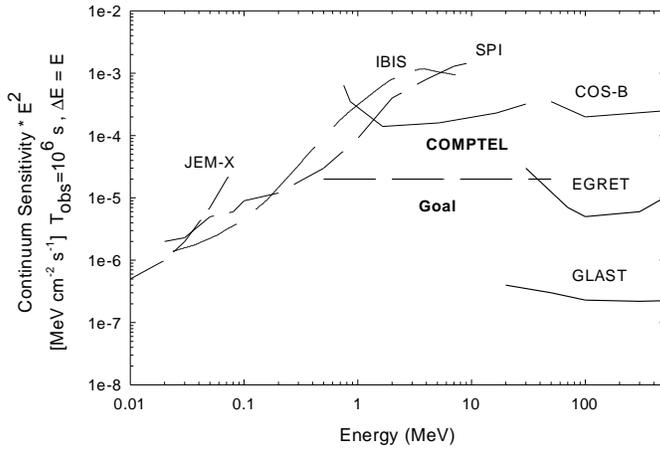


Figure 1: *Sensitivity levels of current and planned γ -ray telescopes (Continuum sources observed for 10^6 s).*

the already scheduled missions in the neighbouring energy bands, we therefore propose and develop a more modest telescope with survey capabilities as a follow up to COMPTEL. Our goal is an improvement of a factor of 10 in sensitivity. Figure 1 shows the sensitivity for past and future gamma-ray instruments and the goal for this next step for a combined Compton- and Pair-creation telescope in the range 0.4 to 50 MeV.

2 Instrument Design and Prototype Development

Combined imaging and spectroscopy of photons with MeV energies is a difficult experimental task. The interaction cross-section at these energies goes through a minimum (masks become transparent) and the primary interaction of MeV photons produces secondary, scattered photons of long range. Around 8 MeV (e.g. in Si) the type of interaction changes from Compton scattering to pair-creation. In order to make full use of all interacting photons in a telescope one requires sufficient material depth to record a good number of events (efficiency), spatial tracking of secondary \sim MeV electrons, and the resolved measurement of all energy deposits. The construction of such instruments is based on solid-state tracking detectors and scintillators (e.g. the TIGRE concept¹⁾), liquid Xenon time projection chambers²⁾ or Germanium detectors³⁾. For MEGA we have chosen the design shown in figure 2. The baseline design for the tracker

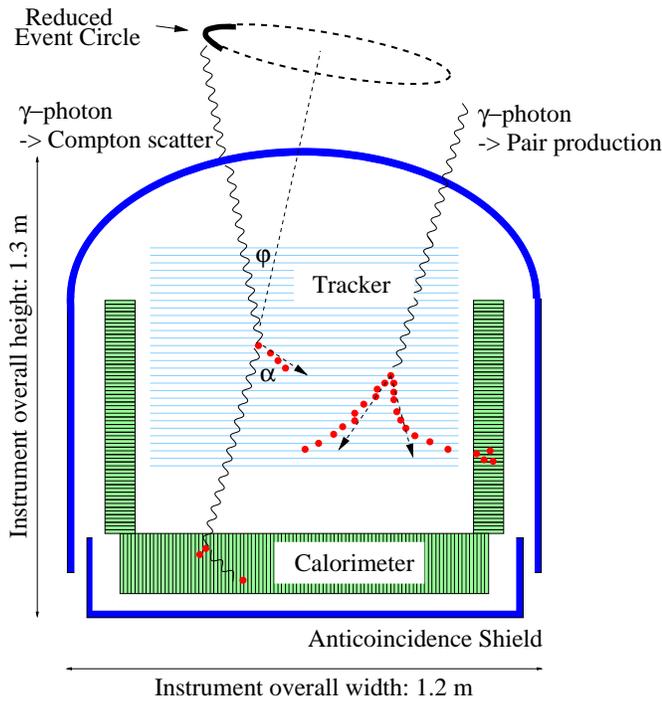


Figure 2: Schematic design of the MEGA detector and typical photon interactions. The tracking of Compton recoil electrons allows to restrict the full event circle to a short arc, which leads to a considerable improvement in the overall sensitivity.

contains 32 layers of double sided Si-strip detectors (6×6 wafers of 6×6 cm² each, thickness $500 \mu\text{m}$, with strip-pitch $\sim 470 \mu\text{m}$). It is enclosed by a calorimeter made of CsI cells ($5 \times 5 \times 40(80) \text{ mm}^3$) read out with Silicon PIN diodes. The detector assembly is surrounded by an anticoincidence shield made of plastic scintillator. A prototype detector for MEGA has been assembled with 10 layers of 3×3 Si wafers and 20 modules of CsI detectors (8, 4, and 2 cm deep with 120 pixels each). Readout of the detectors occurs via an ASIC front-end (TA1 chip by IDE, 128 channels), custom made front-end control units, and laboratory VME electronics. Figure 3 shows the MEGA prototype at a preliminary stage of completion. An actual imaging result using a ^{88}Y source (1836



Figure 3: *The MEGA prototype detector during a preliminary stage of completion. The tracker consists of 10 layers and 4 calorimeter blocks are visible.*

keV) placed above the telescope at a distance of 80 cm is shown in figure 4. Further calibrations with sources and at accelerators are planned for the near future. Simulations for the full MEGA detector based on the presently available prototype properties indicate an effective area of about 100 cm^2 , angular resolution of 2.5 degrees and energy resolution of 3-4% (FWHM at 2 MeV). A very exciting prospect for an instrument based on Compton scattering through large angles is its sensitivity to the polarization of the detected photons. Simulations have been carried out¹⁾ which show that a MEGA like telescope will be sensitive to polarization levels of 10% from a source of Crab intensity after about 100 hours of observation.

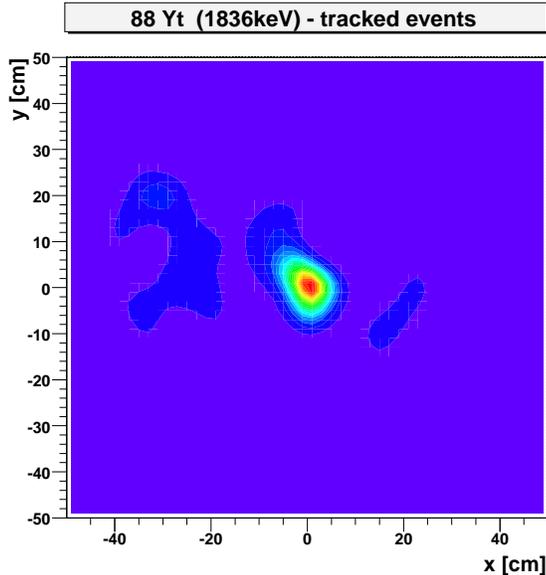


Figure 4: Image of a ^{88}Y source at 1836 keV measured with the MEGA prototype.

3 Mission Concept and Survey Sensitivity

The baseline MEGA detector has been studied for a small satellite mission. The detector will have a mass of about 650 kg and dimensions of 1.3m diameter by 1.1m length. Placed on a standard small satellite platform the launch payload mass is about 950 kg, has a diameter of 2.0 m and 2.3 m length. The electrical power requirement will be ~ 400 W and the average telemetry rate about 50 kbit/s. The development time of MEGA to launch could be about 5 years and an orbital mission of 3-5 years should be foreseen. Operations overlapping with the GLAST mission would be of very high scientific value because of the complementary energy bands of both projects. MEGA should be placed in a low-earth orbit (~ 500 km) close to the equator, to provide an environment with lowest particle background in order to optimize the telescope sensitivity. MEGA is planned to perform continuous all-sky scans with the axis always pointed close to zenith. The large field-of-view describes a wide path of exposure during each orbit and allows to monitor most of the sky continuously for transient sources. Real-time telemetry is planned through the geostationary

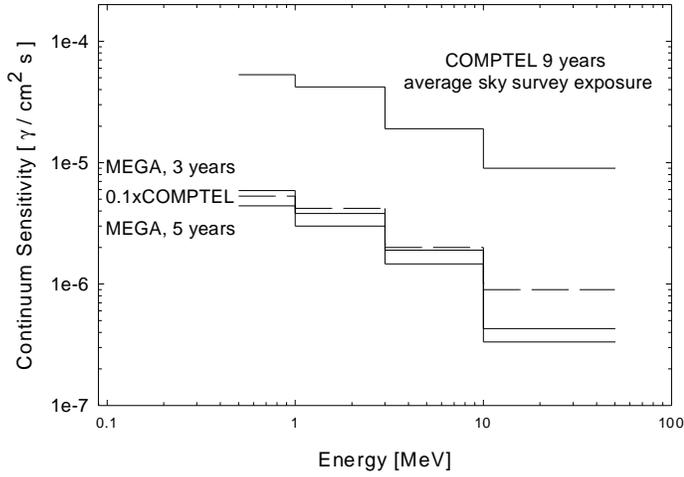


Figure 5: *Sensitivity achievable with MEGA during a 3 or 5 year mission.*

TDRSS relay satellite system. On ground the telemetry data will be analyzed promptly for bursts and transients and appropriate alert messages will be sent out to initiate follow-up observations. We have estimated the survey sensitivity of MEGA with simulations based on GEANT 3. The assumed orbital background for MEGA was taken to be 3 times the COMPTTEL background at 5GV cut-off rigidity. The realism of this assumption will have to be investigated in more detailed background simulations, once the mass and material compositions of MEGA on a satellite and the orbit are defined more accurately. The sensitivities for MEGA in 3 and 5 years of operation compared to the average sensitivity of the COMPTTEL all-sky survey are shown in figure 5.

4 Simulations and Expected Results

Synthetic observational data for MEGA in a scanning survey mode were generated in Monte Carlo simulations. The inversion of these data, which include statistical and systematic errors of energy and spatial resolution, into images is not straightforward and iterative algorithms have to be used. The complex MEGA data space is efficiently handled with the application of novel list-mode

algorithms³⁾. The image of a calibration source in figure 4 was derived with this method.

A preliminary estimate of the number of sources at MeV energies detectable in a 3-year MEGA sky survey predicts that in our galaxy about 100 unidentified EGRET sources should be measured, if a reasonable extrapolation of their high-energy spectra prevails. The number of detected pulsars at MeV energies should grow to about 10 and about a dozen compact galactic binary systems, containing black holes like Cyg X-1, should become visible. The observation of sites of nucleosynthesis, either from explosive events like Novae and Supernovae, or through the radioactive debris of supernova remnants and in the galactic diffuse radioactivity is the other key science objective for MEGA. Due to the continuous survey we expect to detect several highly transient novae each year. Although a prediction is highly model-dependent we expect to discover about 5 young galactic supernova remnants through their radioactive ^{44}Ti line emission. Detailed mapping of the spatial and spectral distribution of the diffuse galactic γ -ray emission, including γ -ray lines from the interstellar medium, will allow to investigate galactic star formation regions, the production of elements in massive stars, and sites of cosmic-ray acceleration.

In extragalactic space about 100 blazars and more than 10 radio and Seyfert galaxies should become visible. Gamma ray bursts will occur and should be imaged in the large field of view of MEGA about once every two days. MEGA should also detect a few supernovae from the nearer galaxies per year. The origin and composition of the diffuse cosmic background at MeV energies is still mysterious and not well measured. At about 5 MeV a transition from predominantly thermal (Seyfert galaxies) to non-thermal ('EGRET Blazars') sources seems to be indicated. However radioactivity from distant supernovae, radio galaxies and new 'MeV' blazars, or clusters of galaxies could also be important new components of the background.

5 Conclusions

The development of a small prototype detector for a next generation medium energy gamma-ray telescope makes good progress. In March 2002 the MEGA prototype will be calibrated at Duke University's accelerator facility providing a High-Intensity Gamma-Ray Source. This installation provides monoenergetic γ -ray photons with tuneable energies between 2 and 55 MeV which are highly polarized. In June 2002 we, in collaboration with CESR Toulouse, plan to fly the prototype detector on a balloon flight from Gap, France. During this balloon flight the background sensitivity and the performance of MEGA will be investigated. If these tests are successful the MEGA concept will be proven suitable for a small satellite mission in the near future.

6 Acknowledgements

The MEGA Collaboration, including teams at CESR, Toulouse (France), UNH, Durham, NASA/GSFC, Greenbelt, NRL, Washington, Columbia University, NY, IGPP-UCR, Riverside and UA, Huntsville (all USA), has provided essential discussions and input to this work. The German Space Agency DLR provided support for the development of the MEGA prototype (50 TT 9738 9) and for a study of the MEGA mission (50 OO 0002).

References

1. O'Neill, T.J. *et al.*, Proc. Fifth Compton Symposium, Portsmouth NH, AIP **510**, 804 (2000).
2. Aprile, E. *et al.* Proc. SPIE, **4140**, 333 (2000).
3. Boggs, S.E. & Jean, P., Astron. Astrophys. Suppl. **145**, 311 (2000).
4. Wilderman, S.J. *et al.* IEEE Trans. Nucl. Sci., **45**, 957 (1998).

Frascati Physics Series Vol. XXIV (2002), pp. 417-420
INTERNATIONAL SCHOOL OF SPACE SCIENCE - 2001 Course on
ASTROPARTICLE AND GAMMA-RAY PHYSICS IN SPACE, L'Aquila (Italy), 30/8 - 7/9, 2001

AN INTRODUCTION TO ASDC, A NEW ASI FACILITY

Sonia Rebecchi

ASI Science Data Center, Consorzio Interuniversitario per la Fisica Spaziale

ABSTRACT

The ASI Science Data Center (as ASDC) is a facility established by the Italian Space Agency (ASI) to act as an interface between ASI satellite missions and the international users community. In the following a brief review of ASDC goals and activities is presented as well an example of the available services.

1 Introduction

A long debated issue over the past years was how to preserve, to make available and to deal with the huge amount of data such as those obtained from space missions. As an example HST and ESO telescopes are producing 4.5 Tbytes per year of scientific data ¹⁾. In order to answer to this need several facilities were created by the main space agencies spread all over the world. See ²⁾ for a list of the available facilities.

With the launch of BeppoSAX in 1996, ³⁾ the Italian Space Agency (ASI) has started to develop the nucleus of a facility to handle, archive the satellite data and distribute them to the international community. From the BeppoSAX experience and also in view of ASI small scientific missions program, which expects to launch a new satellite every 2 years, the ASI Science Data center ASDC ⁴⁾ has been established in November 2000. ASDC web home page can be found at: <http://www.asdc.asi.it/>

2 ASDC current activities

As ASDC is born from the Science Data Center created for BeppoSAX, and a big part of the activities are still related to this mission. Thanks to its unique x-ray broad-band energy range (0.1 – 300KeV) and two different type of instruments onboard, such as narrow and wide field instruments (hereafter NFI and WFC), BeppoSAX proved to be very successful ⁷⁾ and it has performed more than 1700 observations so far.

ASDC current activities for BeppoSAX are:

- Perform a standard analysis for quality check of the NFI data before delivering to the PI
- Re-process old observations with improved new software
- Build a CD archive for both raw data and standard results
- Process all WFC data
- Make all archive levels accessible through the interactive archive
- Provide expert support to the international community by online services and helpdesk.

Other new activities have been added both to extend the number of available catalogues, more than 200 covering all wavelenghts and services. Also of great importance is the ASDC contribution in terms of preparing documentation and reduction software for the SWIFT ⁶⁾ mission to be launched in 2003.

An example of ASDC activity, the interactive archive interface, is shown in Fig. 1). After a query to the interactive archive the software replies by showing all performed BeppoSAX observations as well as possible data rights restrictions.

	Target Name	RA (J2000) hh mm ss.ddegrees	Dec (J2000) dd mm ss.ddegrees	start_time	archive_sta
	↑ ↓	↑ ↓	↑ ↓	↑ ↓	↑ ↓
	GX340+0	16 45 47.5	-45 36 38.9	17-AUG-2000	Restricted to PI
	GX340+0	16 45 47.5	-45 36 38.9	16-AUG-2000	Restricted to PI
	4U0115+63	01 18 31.9	+63 44 21.9	13-AUG-2000	Public access
	X1659-298	17 02 06.2	-29 56 44.9	12-AUG-2000	Public access
	IRASF13305-1739	13 33 16.5	-17 55 00.0	11-AUG-2000	Restricted to PI
	Mrk478	14 42 07.6	+35 26 31.9	09-AUG-2000	Public access

Figure 1: *BeppoSAX* online archive

Then it is possible to ask to display other parameters, to get statistical informations or to get further details on a specific observation and compare BeppoSAX images with those of other x-ray satellites. It is possible to obtain an optical finding chart or perform an own online analysis.

3 Future

Satellites has proved to be very important in many fields and in particular for astronomy where they can provide information at frequencies otherwise unreachable by ground. Therefore scientists will continue to request observations from satellites and the huge amount of data obtained will continuously grow in the future.

As previously mentioned ASI has approved a program of space missions and the first two of them AGILE ^{7, 8)} and DAVID ⁹⁾ will be respectively launched in 2003 and 2004. In collaboration with NASA also ASI will participate in SWIFT a mission mainly intended, thanks to a wide coverage of the sky and to a quick prompt automatic repointing, to study gamma ray bursts. To this and other future missions ASDC facility will address its future challenges.

4 Acknowledgements

The author wishes to thank Dr. Paolo Giommi, for providing this talk figures and a revision of the text, ASI for financial support, Gabriella Ardizzoia, of Consorzio Interuniversitario per la Fisica Spaziale, for her availability, all the BeppoSAX staff and the organizing committee.

5 References

References

1. <http://ecf.hq.eso.org/astrovirtel/>
2. <http://www.cv.nrao.edu/fits/www/yp-center.html>
3. G.Boella *et al*, A&AS, **122**, 299 (1996).
4. P. Giommi, An Overview of the ASI Science Data Center -ASDC- ASI February 2001
5. http://science.msfc.nasa.gov/newhome/headlines/ast12jan98_1.htm
6. P. Caraveo, SWIFT and Gamma ray burst, these proceedings
7. G. Barbiellini, AGILE and the relation between research and technology, these proceedings
8. M. Tavani, The Physics of AGILE, these proceedings
9. <http://ars.rm.asi.it/webars/pmst/pmst.david.html>

GEANT SIMULATION OF THE AGILE GAMMA-RAY IMAGING DETECTOR

Veronica Cocco

Univ. di Roma "Tor Vergata" and INFN, sezione di Roma II, Italy

Francesco Longo

Univ. di Ferrara and INFN, sezione di Ferrara, Italy

Marco Tavani

Istituto di Fisica Cosmica, CNR, Milano, Italy

ABSTRACT

Using a GEANT-based Simulator of the Gamma-Ray Imaging Detector (GRID) developed for the AGILE space astrophysics mission, we optimized the event trigger processing. In this paper, we describe the AGILE instrument geometry and the model assumed for the charged particle and the albedo-photon backgrounds. We present the main results on different levels of data processing and obtain the background rejection efficiency of the GRID detector.

1 Introduction

AGILE is an ASI Small Scientific Mission dedicated to high energy astrophysics ¹). The AGILE instrument is designed to detect and image photons in the 30 MeV - 50 GeV and 10 - 40 keV energy bands, with excellent spatial

resolution and timing capability and an unprecedentedly large field of view covering $\sim 1/5$ of the entire sky at energies above 30 MeV. The AGILE gamma-ray mission requires a low-background orbit to maximize its scientific output.

The optimization of the AGILE design was obtained through a Monte-carlo study of the detector performance, using the GEANT 3.21 code ²⁾. In this paper, we describe the AGILE instrument and the particle/albedo-photon background models assumed for the optimization of the on-board data processing. We outline the adopted trigger strategies and present the main results about the on-board background rejection.

2 The AGILE instrument model

The AGILE scientific instrument is made of three integrated detectors with broad-band detection and imaging capabilities ^{1, 3)}. The AGILE Gamma-Ray Imaging Detector (GRID) consists of a Silicon-Tungsten Tracker, a Cesium Iodide Mini-Calorimeter, an Anticoincidence system made of segmented plastic scintillators, fast readout electronics and processing units. The Super-AGILE detector will provide detection and imaging capabilities in the hard X-ray range. The CsI Mini-Calorimeter will also detect and collect events independently from the GRID in case of impulsive transients.

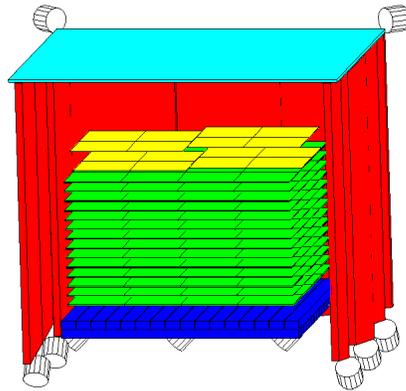


Figure 1: *A simplified view of the AGILE instrument model.*

In the simulation code we modelled the GRID detector according to the AGILE design ³⁾. Fig. 1 shows a simplified view of the AGILE instrument model used in our simulations.

3 Background assumptions

A quasi-equatorial orbit is preferred for the AGILE mission and will provide a relatively low-background environment. Taking into account data from SAS-2 and Beppo-SAX missions, we expect an average rate of charged particle background above ~ 1 MeV of ~ 0.3 particles $\text{cm}^{-2} \text{s}^{-1}$ for a quasi-equatorial orbit near 550 km. The charged particle energy spectra assumed in our simulations are shown in Fig. 2. They are based on data from the 1998 AMS Shuttle flight ^{4, 5}), and from the MARYA experiment on board of the MIR space station ⁶). The interaction of the charged cosmic-rays with the upper atmosphere

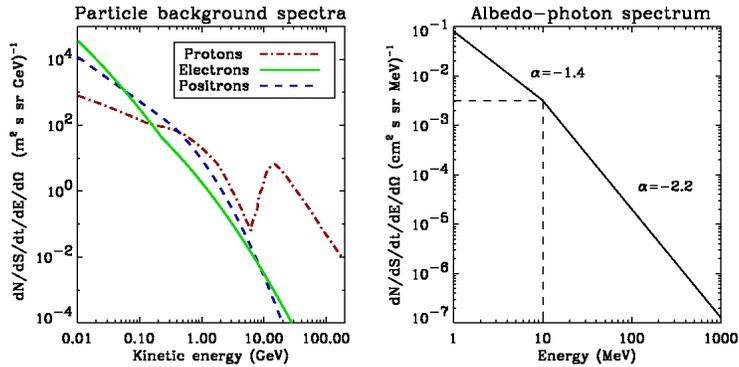


Figure 2: *Left Panel: Charged particle background energetic spectra assumed in our simulations. Right Panel: Average albedo-photon energy spectrum.*

induces a relatively strong gamma-ray background peaking at the Earth horizon. This effect involves a localized increase of the gamma-ray emission that we properly took into account on the basis of SAS-2 ⁷) and balloon data ⁸). Fig. 2 shows the average flux of albedo photons over the solid angle of the subtended Earth surface at the height of 550 km.

4 Trigger Strategies and Background Rejection

We studied different GRID trigger configurations, and optimized their performance. The baseline GRID trigger logic consists of two different levels. A (hardware) Level-1 trigger logic uses the information from the Silicon detec-

tors and AC panels and considers also a simplified view of the event topology obtained by the front-end chips. Level-1 trigger reduces the charged-particle background from a rate of ~ 2000 Hz to a rate of ~ 60 Hz. A (software) Level-2 on-board data processing makes a crucial use of the analog (charge) information in the Si-microstrips for a refined view of the event topology at the “cluster” level. Level-2 on-board processing also selects events based on a simplified photon direction reconstruction (necessary to reject Earth albedo photons). After the on-board Level-2 processing, we can reduce the total (charged particle and albedo-photon) background rate to $\sim 20-30$ Hz, without affecting significantly the cosmic gamma-ray detection.

References

1. Tavani M. et al., to be published in the Proceedings of Gamma2001 Conference
2. Giani S. et al., CERN Long Writeup W5013 (1994)
3. Barbiellini G. et al., to be published in the Proceedings of Gamma2001 Conference
4. Alcaraz J. et al., Physics Letters B 472, 215 (2000)
5. Alcaraz J. et al., Physics Letters B 484,10 (2000)
6. Koldashov S.W. et al., 24th ICRC 4, 993, (1995)
7. Thompson D.J., Simpson G.A. and Özel M.E., Journal of Geophysical Research 86, 1265 (1981)
8. Costa E., Massaro E., Salvati M. and Appolloni A., Astrophysics and Space Science 100, 165 (1984)
9. Cocco V., Longo F. and Tavani M. (2001), in press on GAMMA 2001: Gamma-Ray Astrophysics 2001, Ritz S., Gehrels N. and Shrader C.R.(eds), AIP Conference Proceedings 587
10. Longo F., Cocco V. and Tavani M. (2001), NIM-A submitted
11. Cocco V., Longo F. and Tavani M. (2001), NIM-A submitted

**PRESENTATION OF THE AGILE RECONSTRUCTION
METHOD AND KALMAN FILTER ALGORITHMS FOR
GAMMA-RAY SILICON DETECTORS IN SPACE**

Carlotta Pittori

*Dip. di Fisica, Univ. di Roma "Tor Vergata" and INFN
Sez. di Roma II, I-00133 Roma, Italy*

Andrea Giuliani, Sandro Mereghetti and Marco Tavani
IFC-CNR, V. Bassini 15, 20133 Milano - Italy

ABSTRACT

The AGILE REconstruction Method (AREM) is a method of γ -ray direction reconstruction to be applied to high-resolution Silicon Tracker detectors in space. AREM correctly addresses three points of the analysis relevant for off-axis incidence angles: 1) intrinsic ambiguity in the identification of the 3-D tracks and conversion plane; 2) proper identification of the 3-D reconstructed γ direction; 3) careful choice of an energy weighting scheme for the 3-D tracks. The excellent spatial resolution obtained by the AGILE Silicon Tracker allows to improve the angular resolution by a factor ~ 2 at energies $\gtrsim 400$ MeV with respect to previous spark chamber detectors (e.g. EGRET).

1 Introduction

AGILE (Astro-rivelatore Gamma a Immagini LEggero) is a Small Scientific Mission of ASI (Agenzia Spaziale Italiana) with a γ -ray imaging system based

on state-of-the-art Silicon strip technology ¹⁾ and it will have an unprecedentedly large field of view, ~ 3 sr, and a very good intrinsic spatial resolution. CERN testbeams show that, by using the analog readout, which gives information on the charge distribution in Si-microstrips, one can achieve a spatial resolution of order of $\sim 40 \mu\text{m}$ ²⁾. Despite the relatively small dimensions and effective area, the AGILE goal is to obtain the best sensitivity ever reached for off-axis events, preserving a on-axis sensitivity comparable to that of EGRET. Therefore, the angular resolution optimization becomes a crucial point to fulfil the mission scientific objectives.

2 The AREM method

The γ -ray direction reconstruction is based on the process of pair production, and is obtained from the analysis of the e^+/e^- tracks originating from a common vertex. The current customary “2-D projection method” of analyzing separately the two tracks projections in the ZX and ZY Tracker views induces two kinds of systematic error in the photon direction reconstruction: A) the intrinsic ambiguity in the proper identification of the two 3-D tracks; B) the problem of the identification of the true 3-D reconstructed γ direction. Finally, we emphasize the importance of the: C) choice of track weighting scheme. In general, the photon energy is not evenly divided between the two particles. Since the direction of the most energetic particle is closer to that of the incident photon, an “energy-weighted” reconstructed direction should be computed¹.

As for the point A), we note that when the e^+/e^- pair hits simultaneously the active Tracker layers, the signal will correspond to two projected track points in each ZX and ZY view, but it could correspond to two possible couples of points in space, as shown in Fig. 1-A. This gives rise to an ambiguity in the conversion plane identification. The “projection problem”, point B), stems from the fact that in previous γ -ray experiments, after the identification of the two projected tracks in each view, the next step was to take their (eventually weighted) bisectors and compose them to obtain the reconstructed γ -ray direction. As illustrated in Fig. 1-B, the true 3-D bisector is different from the one obtained from the two bisecting lines in each projected view. As shown in

¹For simplicity, in the following we illustrate points A) and B) in case of an even energy sharing, i.e., when the photon direction coincides with the bisector.

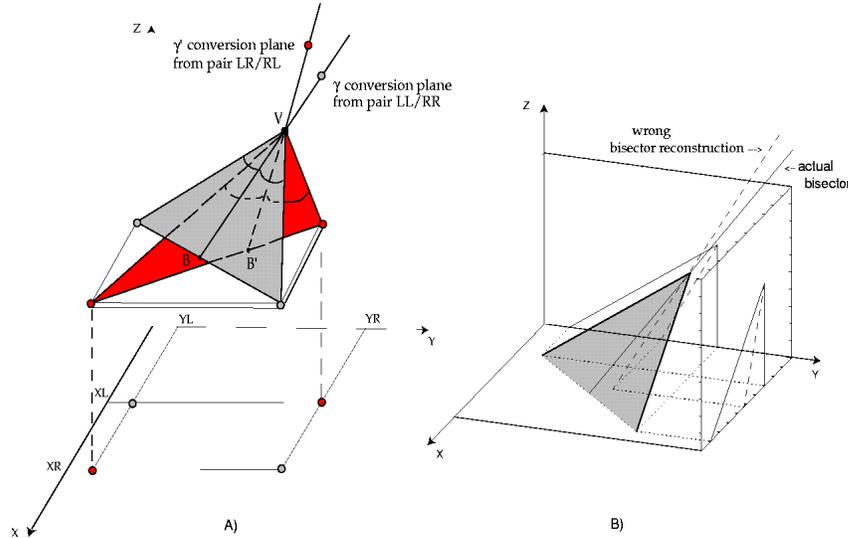


Figure 1: A) The “conversion plane problem”: make the right choice for one out of two possible conversion planes in 3-D. B) The “projection problem”: the projections of the bisector are different from the bisectors of the projections.

ref. 3), this systematic effect increases for increasing off-axis angles and large opening angles, up to values of $\sim 0.5^\circ$, and it would have a significant impact for AGILE. Furthermore, with high resolution Si-detectors it is possible to estimate the e^+/e^- energies from a few MeV to the GeV scale, by measuring deviations due to multiple scattering. This fact allows to properly define the weight of each track for the direction reconstruction (point C).

As described in detail in ref. 4), AREM is a 3-D reconstruction method, which takes into account these three points of the analysis. Several approaches are under study: (i) first n-Planes Resolution, using only information from the first hit planes (2PR, 3PR, ...), (ii) algorithms based on the Kalman filter 5) for an optimal use of the information from all hit planes. The Monte Carlo simulations of the AGILE-GRID imaging performance were done using the GEANT 3.21 code 6). In Fig. 2 we show, as an example, the 3-D PSF obtained by using only information from the first 3 hit Tracker planes (3PR) for near-on-axis events at $E_\gamma = 1$ GeV. In Fig. 3 we show the 3-D PSF distribution profiles obtained from the AGILE Kalman filters algorithms (AKF) for several angles and energy values. The 3PR provides a satisfactory PSF for near on-axis events,

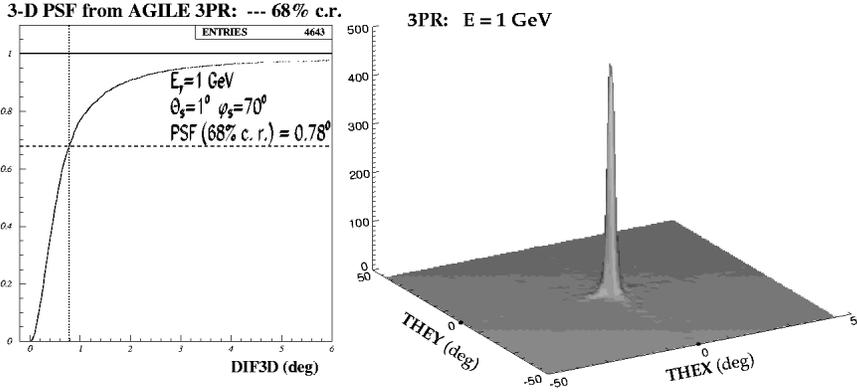


Figure 2: *AGILE* on-axis 3-D PSF from the 3PR reconstruction for $E_\gamma = 1$ GeV. The left panel curve represents the integral distribution of the difference between true and reconstructed direction of each photon.

compatible with AKF, even though with a lower reconstruction efficiency. The AKF provides a good event reconstruction with very high efficiency (above 90% for $E_\gamma > 200$ MeV) for a variety of incidence angles. Finally, in Fig. 4 we compare the preliminary *AGILE* angular resolution, between 0° and 50° off-axis, with that of *EGRET* on-axis. The figure shows the 3-D 68% containment radius as a function of energy. The *AGILE* 3-D PSF is better than that of *EGRET* by a factor of ~ 2 above 400 MeV.

References

1. M. Tavani *et al.*, AIP Conf. Proc. (ed. M. McConnell), **510**, 746, (2001)
2. G. Barbiellini *et al.*, Proc. of the GAMMA 2001 symposium, to be published by AIP, and preprint (2001) submitted to NIM.
3. A. Giuliani, Laurea Dissertation, Univ. di Pavia (2001).
4. C. Pittori and M. Tavani, ROM-2F/2001/12, submitted to NIM.
5. R. Frühwirth, NIM A, **262**, 444 (1987).
6. V. Cocco, F. Longo and M. Tavani, *AGILE-SIM-TN-001*, Issue n.2 (2000), and preprint submitted to NIM.

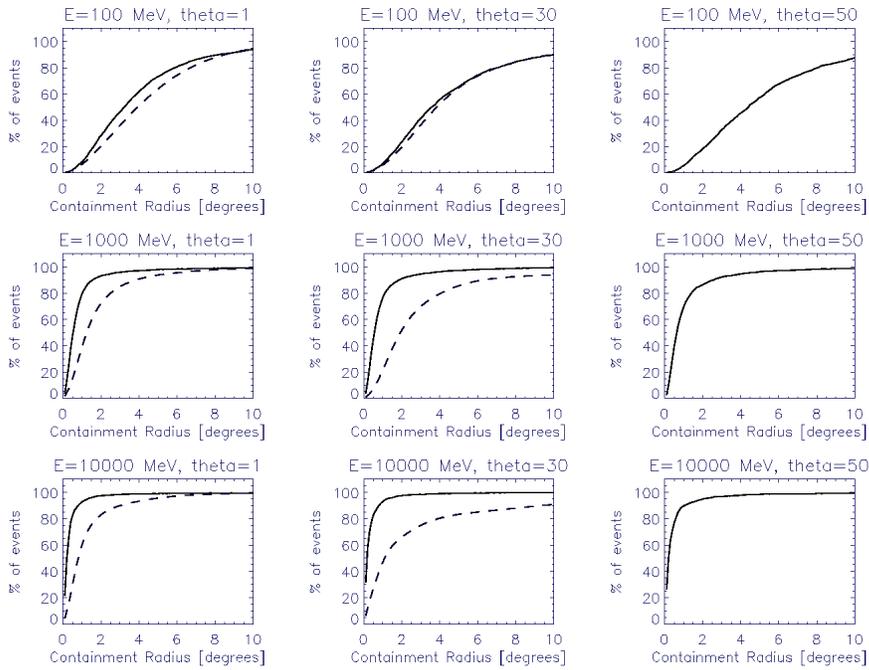


Figure 3: *3-D integral PSF profiles obtained with the AGILE-GRID Kalman filters algorithms (solid curve) compared to the corresponding EGRET values (dashed curve) when available (public data).*

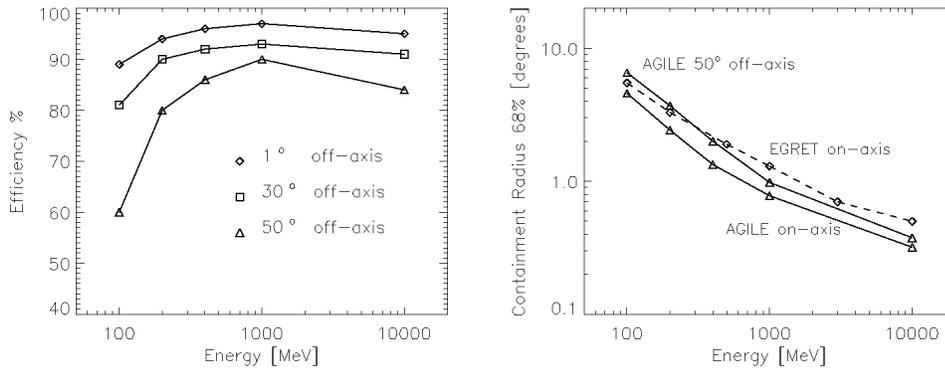


Figure 4: *Left panel: Reconstruction efficiency of the AGILE Kalman filters algorithms for different off-axis directions. Right panel: Preliminary results for the AGILE 3-D containment radius (68%).*

The dark side of the Universe

P. De Bernardis	The density of the Universe
J.Primack	The Nature of Dark Matter
P.Ullo	Exotic cosmic rays from pair annihilations of dark matter wimps
A.Lionetto	MSSM parameters space regions allowed by indirect neutralino detection

THE DENSITY OF THE UNIVERSE

Paolo de Bernardis and Silvia Masi
Dipartimento di Fisica, Universita' La Sapienza, Roma

ABSTRACT

The significance and the measurement of the average density of the Universe is reviewed at introductory level. Current observations of the Cosmic Microwave Background anisotropy measure the density to be critical with an accuracy of the order of 10%.

1 Which density ?

Cosmology studies the Universe at large scales (> 100 Mpc). Building blocks for the visible Universe are the Galaxies. The galaxy distribution is close to homogenous at very large scales (see e.g. ¹⁾ ²⁾). So, as a zero order approximation, Cosmology studies the evolution of a homogeneous medium.

The only force effective at very large scales is gravity, and we believe that the correct theory of gravity is general relativity. For this reason, we will refer

to the mass-energy density of the idealized homogeneous medium when we will speak about the density of the Universe. We need to include all the forms of energy contributing to the stress-energy tensor in Einstein's field equations.

Note that the assumption of a homogenous model is crucial. The density of a fractal distribution of matter, for example, is not even defined.

2 Mass-Energy Density and the Expansion of the Universe

The average mass-energy density is a very important parameter. We know, since Hubble's observations, that the Universe is expanding now. The expansion rate is described by the Hubble constant $H_o \sim 70 \text{ km/s/Mpc}$ ³⁾. The Friedmann model of the Universe describes such a system, solving Einstein's field equation in the isotropic homogeneous case ¹³⁾. The striking result of this General Relativistic analysis is that a static Universe is forbidden. Assuming that only matter and radiation are present, there are three families of solutions: expanding forever, critical, expanding and re-collapsing. The kind of expansion depends on the amount of mass-energy present in the Universe. The critical case, where expansion decelerates to complete rest only for $t \rightarrow \infty$, happens when the kinetic energy of the expansion equals the gravitational energy. We can easily find out the mass-energy density ρ_c required in this case. Consider one galaxy, mass m , at distance r from the origin O. The galaxy will recede with speed $H_o r$ and will feel the gravity of all the matter (mass M) inside the sphere of radius r (the Birkhoff theorem of General Relativity is used here). We have

$$\frac{1}{2}mv^2 = \frac{GMm}{r} \quad (1)$$

and using H_o and the density ρ_c :

$$\frac{1}{2}mH_o^2r^2 = \frac{G\frac{4}{3}\pi r^3\rho_c m}{r} \quad (2)$$

so that

$$\rho_c = \frac{3H_o^2}{8\pi G} \quad (3)$$

This is the critical density, numerically $1.88 \times 10^{-26} h^2 \text{ Kg/m}^3$, where $h = H_o/(100\text{km/s/Mpc})$. If the mass-energy density ρ of the Universe is larger than ρ_c , the expansion decelerates to zero and the Universe collapses after a finite time. If $\rho < \rho_c$, the Universe expands forever. The ratio $\Omega = \rho/\rho_c$ is

called the density parameter. The density of the Universe is decreasing during expansion, and at early times all the solutions come from a state of infinite density, the so called Big-Bang.

There is a paradox in this family of solutions. We measure today, some billion years after the big bang, a value of ρ of the order of ρ_c . This means that in the past ρ had to be incredibly close to ρ_c , otherwise we would not exist, either because the Universe would have already collapsed or because it would be already completely diluted. This is a problem of fine tuning of the initial conditions, well known as the *paradox of flatness*, which cannot be solved in the framework of the standard Hot Big Bang theory. The theory of cosmic inflation in the very early Universe offers a solution to this problem and to several others ⁴⁾.

We live in the present epoch, and we only know the present expansion rate of the Universe, measured by the Hubble constant. To know more about its previous and future evolution we need to measure Ω . Moreover, we need to understand what contributes to it. Let's list all the possible contributions.

In the Universe there is light and there is electromagnetic radiation. A lot of photons (about $400 \gamma/\text{cm}^3$). And a lot of neutrinos. But photons and neutrinos today have a mass-energy density which is negligible with respect to the critical density.

We see stars and interstellar gas and dust. Stars are grouped in galaxies which are grouped in galaxy clusters. Hot gas fills the inter-cluster volume. All this luminous (baryonic) matter accounts at most for about 5% of the critical density: $\Omega_b < 0.05$ (see ⁵⁾ for a review).

We know that there is dark matter, which is needed to explain the dynamics of stars in Galaxies and of galaxies in clusters. This dark matter does not interact with light, and we do not know which particles is made of (cold baryons ? neutrinos ? WIMPS, neutralinos ?) Its density is also largely unknown, but estimates are converging to 20-30% of the critical density ($\Omega_m = \Omega_{DM} + \Omega_b \sim 0.3$) ^{6), 7), 8)}.

An additional possibility is that in the universe there is some quantity of repulsive, dark energy: a form of energy with negative pressure (vacuum energy, or cosmological constant Λ , or quintessence). $\Omega_\Lambda \sim 0.7$ would account for the recent evidence for an accelerated expansion of the universe, coming from the observations of distant supernovae (^{10), 11)}). The possible presence of "dark

energy” is quite important for the evolution of the Universe. At some point, ”dark energy” becomes the dominant form of energy, driving an accelerated expansion even if the curvature is positive. If dark energy turns out to be present, curvature and destiny of the Universe are not related anymore (12).

In summary, our knowledge of the content of the Universe is very poor, and we measure directly only a very little fraction of it. We need a different approach to measure Ω .

3 Mass-Energy Density and the Curvature of the Universe

An independent method is the determination of the geometry of the Universe, by means of the observation of its effects on light propagation. According to General Relativity, the presence of mass and energy curves the space. This was experimentally demonstrated since 1919, by observation of the deflection of light coming from distant stars, and passing close to the mass of the Sun. The same phenomenon is now usual in the images of distant sources by the Hubble Space Telescope 9). These sources are deformed by the presence of mass (galaxies or clusters) on the light path, and are often seen as multiple images of the same source (Gravitational lensing).

The large scale geometry of the Universe is controlled by the average mass-energy density in the same way: its presence curves the background metric of the universe. If $\Omega = 1$, the geometry is the standard Euclidean geometry, the space is not curved, and light rays starting parallel will travel parallel forever. If $\Omega > 1$, the additional mass-energy will curve the space so that the global curvature is positive. Two light rays starting parallel will converge. If $\Omega < 1$, the defect of mass-energy will curve the space so that the global curvature is negative. Two light rays starting parallel will diverge.

We need a method to detect the curvature of light rays due to the background metric of our Universe. It is a small effect locally, so we need to perform the test over cosmological scales. If we could place a standard ruler at very large distance, we could do the test by measuring the angle subtended by the ruler. For $\Omega = 1$ the angle would be the one computed from elementary geometry in an Euclidean space. For $\Omega > 1$, we expect to see a larger angle, because light rays converge towards the observer. For $\Omega < 1$ we expect to see a smaller angle.

The angular size of a source with proper linear size d is $\theta = d/d_A$. In

the Friedmann-Robertson-Walker metric, the angular-diameter distance d_A as a function of the redshift z and of the different density parameters can be derived in closed form (14):

$$d_A = \frac{1}{\Omega_k^{1/2} H_o (1+z)} \sinh \left[\Omega_k^{1/2} \int_{\frac{1}{1+z}}^1 \frac{dx}{\sqrt{\Omega_\Lambda x^4 + \Omega_k x^2 + \Omega_M x}} \right] \quad (4)$$

Here $\Omega_k = 1 - \Omega_M - \Omega_\Lambda = 1 - \Omega$; the sinh function must be used for $\Omega < 1$ and has to be replaced with a sin function for $\Omega > 1$. In principle one could use the diameters of Galaxies as a standard rulers, and infer the Ω s from a best fit to $d_A(z)$. But the diameter of Galaxies is hard to define, and evolution effects are important. High redshift Galaxies tend to be more irregular. Similar evolution problems are present for radio-galaxies. The solution to this problem came recently from the study of the first high resolution images of the Cosmic Microwave Background.

4 The Image of the Cosmic Microwave Background

Since the Universe is expanding, it has been denser and hotter in the past. We have a way to investigate directly early epochs, due to the finite speed of light: looking far away is the same as looking in the past. If we look back far enough, we will see an epoch when the universe was about as hot as the surface of the Sun, at a temperature of several thousand K. Then the universe was a hot plasma, called "the primeval fireball", where photons and matter were in thermal equilibrium, due to Thomson scattering of photons against free electrons. We cannot see earlier epochs - as we cannot see inside the Sun - because the plasma is opaque to light. We expect to see light coming from there. Due to the expansion of the Universe, that visible and near infrared light has been red-shifted by a factor 1000, and is now a faint glow of microwaves: the Cosmic Microwave Background (CMB). This has been detected as a 2.73 K blackbody (16). In the framework of the Hot Big Bang model, when we look to the Cosmic Microwave Background we look back in time to the epoch when temperature decreased below ~ 3000 K for the first time. This is the end of the plasma era, at a redshift ~ 1000 , when the universe was ~ 50000 times younger, ~ 1000 times hotter and $\sim 10^9$ times denser than today. It is a transition epoch (called "recombination"): the ionized, opaque Universe becomes neutral and transparent to light, so that CMB photons can travel freely all the way to our

telescopes . The other very important consequence of the transition regards the evolution of density perturbations ($\Delta\rho/\rho$). Density perturbations were oscillating in the primeval fireball as a result of the opposite effects of the pressure of photons and of gravity. After recombination, photons pressure becomes unimportant, and $\Delta\rho/\rho$ can grow and create, through gravitational instability, the hierarchy of structures we see today in the nearby Universe. There are three physical processes converting the density perturbations $\Delta\rho/\rho$ present at recombination into *observable* CMB temperature fluctuations $\Delta T/T$. They are: the photon density fluctuations δ_γ , which can be related to the matter density fluctuations $\Delta\rho$ once the kind of perturbations is specified; the gravitational redshift of photons scattered in an overdensity or an underdensity with gravitational potential difference ϕ_r (Sachs-Wolfe effect ¹⁵); and the Doppler effect produced by the proper motion with velocity v of the electrons scattering the CMB photons. In formulas:

$$\frac{\Delta T}{T}(\vec{n}) \approx \frac{1}{4}\delta_{\gamma r} + \frac{1}{3}\frac{\phi_r}{c^2} - \vec{n}\frac{\vec{v}_r}{c} \quad (5)$$

where n is the line of sight vector and the subscript r labels quantities at recombination. If we are able to measure an image of the CMB, we can estimate the density fluctuations present in the Universe at recombination, which are the early progenitors of the structures present today in the Universe. Moreover, the density fluctuations present at recombination result from processes in the very early Universe, and the observation of the CMB seems to be a good way to investigate them.

5 A standard ruler at recombination

The fact of interest here is that there is a *characteristic scale* in the CMB anisotropy at recombination, because there is an *acoustic horizon* at recombination. The acoustic horizon is determined by the distance sound waves can travel since the Big Bang. The speed of sound in the primeval plasma is close to the speed of light (basically $c_s = c/\sqrt{3}$). The size of the acoustic horizon increases with time. Density perturbations of a given proper size start to oscillate only when the acoustic horizon encompasses them. As long as they are larger than the acoustic horizon, they just follow the expansion of a slightly over-dense or under-dense universe, and do not oscillate. When the acoustic

horizon becomes larger than their size, the opposite edges of the perturbation begin to feel the effects of the forces, and the acoustic oscillation starts. The size of the horizon at recombination separates perturbations which have never oscillated from those that have undergone acoustic oscillations. Let's focus on perturbations with a size slightly smaller than the acoustic horizon. These perturbations will start the oscillation just before recombination, and will have just enough time to fully compress. They will be evident as slightly hotter or slightly colder spots in the image of the CMB. The age of the Universe at recombination is of the order of 3×10^5 years, so the characteristic proper size of these spots is 3×10^5 light-years. This is the standard ruler we needed at cosmological distances. Perturbations slightly smaller or slightly larger than these will not be fully compressed or rarefied, and will not be as evident as these in the image of the CMB. The distance between us and recombination is of the order of the present age of the Universe, i.e. $\sim 15 \times 10^9$ light years. In an Euclidean Universe, the apparent size of these spots will be simply $\theta \sim 10^3 \times 3 \times 10^5 / 15 \times 10^9$ rad, i.e. about one degree (the factor 10^3 comes from the expansion of the Universe after recombination). If, instead, the geometry of the Universe is not Euclidean ($\Omega \neq 1$), then we expect to see magnified spots ($\theta > 1^\circ$) if $\Omega > 1$, and de-magnified spots ($\theta < 1^\circ$) if $\Omega < 1$. The recipe for this measurement is thus:

- Take a picture of the CMB with sufficient angular resolution, $\Delta T(\alpha, \delta)$.
- Compute the angular power spectrum of the image, c_ℓ . Here $c_\ell = \langle |a_{\ell, m}|^2 \rangle$, where $\Delta T(\alpha, \delta) = \sum_{\ell, m} a_{\ell, m} Y_m^\ell(\alpha, \delta)$.
- Check if there is a peak at multipoles corresponding to an angular scale of the order of 1° ($\ell \sim \pi/\theta \sim 200$). The value of Ω will be determined by the location of the peak (i.e. the typical projected angular size of the spots).

The qualitative analysis above can be made rigorously (see e.g. 17), 18), 19). The result is that the location of the first peak in the angular power spectrum of the CMB is indeed a strong function of Ω , but also of Ω_M (or Ω_Λ) 18). In the region $\Omega \sim 1$, it results $\ell_p \sim \Omega^{-\nu}$, with $\nu \sim 2.2, 1.6, 1.25, 0.4$ if $\Omega_M \sim 0.2, 0.3, 0.4, 1$.

There is a lot more information encoded in the angular power spectrum of the CMB (see 20) for a recent review). Multiple peaks are present in the spectrum. The relative amplitudes are sensitive to the physical density of baryons $\Omega_b h^2$, and to the initial power spectrum of the density fluctuations.

This is usually parametrized as $\Delta\rho/\rho(k) \sim k^n$ (here k is the wavenumber, i.e. the inverse of the physical wavelength of the perturbation). Inflationary models ⁴⁾ naturally produce adiabatic, gaussian density fluctuations, with a power spectrum with $n \sim 1$. To a lesser extent, the angular power spectrum of the CMB is also sensitive to other cosmological parameters, like the reionization optical depth τ_C , the different mass-energy densities, and the Hubble constant. The main purpose of the space missions MAP ²¹⁾ and Planck ²²⁾ is the measurement of the power spectrum of the CMB with sufficient resolution and accuracy to allow a precision measurement of all the cosmological parameters.

6 Measuring the Image of the Cosmic Microwave Background

The hunt for detecting anisotropies in the CMB and possibly build an image of it started in the 70s. The CMB turns out to be extraordinarily isotropic: apart from the dipole component due to the motion of the Earth, it became clear soon that any intrinsic anisotropy had to be smaller than 100 parts per million. Measurements are extremely difficult, due to the small amplitude of the anisotropy and to the variability of the atmospheric emission, which is much larger than the signal to be measured. Measurements must be carried out either in the microwaves, from cold, dry sites, or in the far infrared, outside the earth atmosphere, using stratospheric balloons, rockets and satellites. After 27 years of pioneering efforts, the COBE satellite of NASA produced the first detection in 1992. The DMR instrument aboard of the COBE satellite measured the first map of CMB anisotropy ²³⁾. Due to the coarse angular resolution of the instrument ($\sim 7^\circ$ FWHM), only large scale structures could be detected. The contrast of the resulting image is very low, but there are structures, at a level of 10 ppm: $\Delta T_{rms} = (30 \pm 3)\mu K$. The consequence of this observation is that the early universe is smooth at large scales. This is not trivial, because of causal horizon effects: regions of the sky distant more than $\sim 1^\circ$ are not in causal contact at recombination. Why regions many degrees apart feature the same temperature, within a few tens of ppm? This is called the paradox of horizons, and has been solved only with the inflation theory. The angular power spectrum of the CMB was measured for $\ell < 20$ and is consistent with the Harrison-Zeldovich power spectrum $P(k) = Ak^n$ with $n = (1.0 \pm 0.3)$ ²⁴⁾. This is consistent with the predictions of the inflation theory. The resolution of COBE was not enough to make the measurement of Ω outlined above, which requires

an angular resolution better than 1° FWHM. In a long lasting experimental effort, the degree-scale and sub-degree-scale anisotropy has been detected by several ground based and balloon-borne experiments. Only recently, however, it has been possible to detect the presence of peaks in the power spectrum (25), (26), (27), and to produce images where the sub-degree anisotropy is clearly visible (28), (29), (30), (31).

Let's focus now on the BOOMERanG experiment, which produced the first wide sky maps where sub-horizon structures in the CMB are detected with high confidence (28). The instrument setup has been described in detail in (35), (36), (37). It is a scanning telescope, featuring important improvements with respect to previous experiments. First, BOOMERanG uses a long duration (7 to 14 days) stratospheric (~ 38 km) balloon flight around Antarctica. Long integrations on a wide sky region are obtained, along with careful and extensive tests for systematic effects. In addition, flying during the austral summer from Antarctica, the lowest Galactic contamination region of the sky (38) is visible at an azimuth almost perfectly opposite to the azimuth of the Sun. This simplifies the necessary shielding required for thermal and optical reasons. Second, BOOMERanG uses a very sensitive total power receiver, based on spider-web bolometers (39) cooled to 0.28 K with a custom cryogenic system (40), (41). The power detected from one direction is compared to the power from contiguous directions by slowly scanning the telescope ($1^\circ/s$ to $2^\circ/s$ in azimuth). This strategy is enabled by the intrinsic stability of the readout electronics (a low noise AC bridge) and of the detectors. The full payload is gently moved, avoiding mechanical choppers and the related inefficiencies and slowly varying offsets. Third, the focal plane is multiband, with 8 pixels and 4 colors (90, 150, 240, 410 GHz) strategically located with respect to the scan direction in order to have several temporal and spectral confirmations of the detected structures. We track the azimuth of the best sky region while scanning at constant elevation, thus obtaining highly cross-linked maps. BOOMERanG has obtained maps of about 3% of the sky at the four frequencies, with angular resolution ranging between 18 and 12 arcmin for the different channels. The structure evident in the maps at 90, 150 and 240 GHz is spectrally consistent with CMB anisotropy, strongly excluding a local origin of the detected temperature fluctuations (42). The power spectrum has been estimated from the central part of the map, corresponding to about 1% of the sky in (28) and 1.8% of the sky in

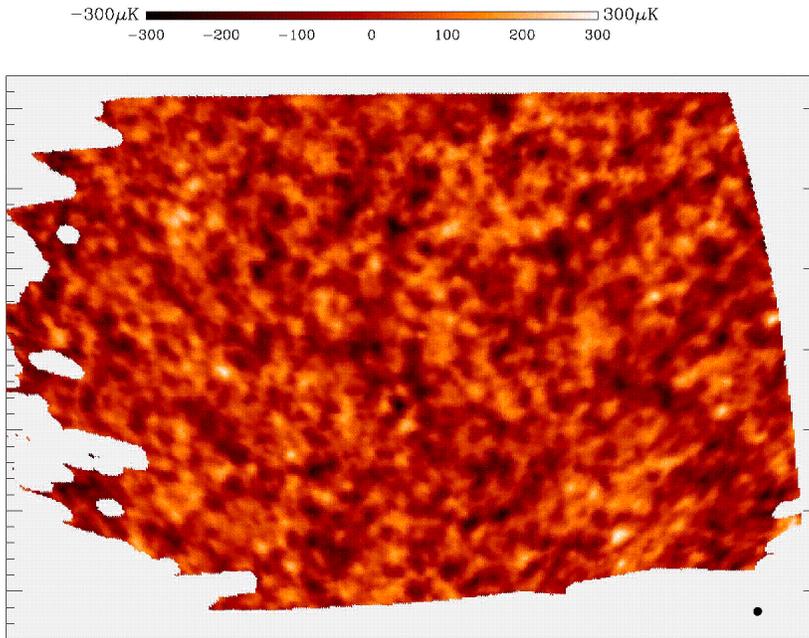


Figure 1: *Map of about 3% of the sky obtained at 150 GHz from one of the BOOMERanG detectors. The black disk has a diameter of 0.5° . The structures dominating the figure are temperature fluctuations of the Cosmic Microwave Background due to acoustic oscillations in the primeval plasma. The rms amplitude of the fluctuations is $\sim 80\mu K$, while the characteristic size of the hot and cold spots is $\sim 1^\circ$.*

30). The map derived from one 150 GHz channel is visible in fig.1. Multiple peaks are evident in the angular power spectrum ³²⁾, the first one located at $\ell \sim 210$. The presence of multiple peaks is the confirmation of the presence of acoustic oscillations in the plasma before recombination, and allows the detection of several important cosmological parameters ³²⁾, ³³⁾, ³⁴⁾. Ω is detected by measuring the location of the first peak. The angular power spectrum of the CMB detected by BOOMERanG ²⁸⁾, ³⁰⁾, DASI ³¹⁾ and MAXIMA ²⁹⁾, ³¹⁾ is plotted in fig.1. Given the orthogonality of the experimental and analysis methods, the agreement of the three results is very good, at least visually. The existing anti-correlations in the bandpowers, and the presence of some overlap

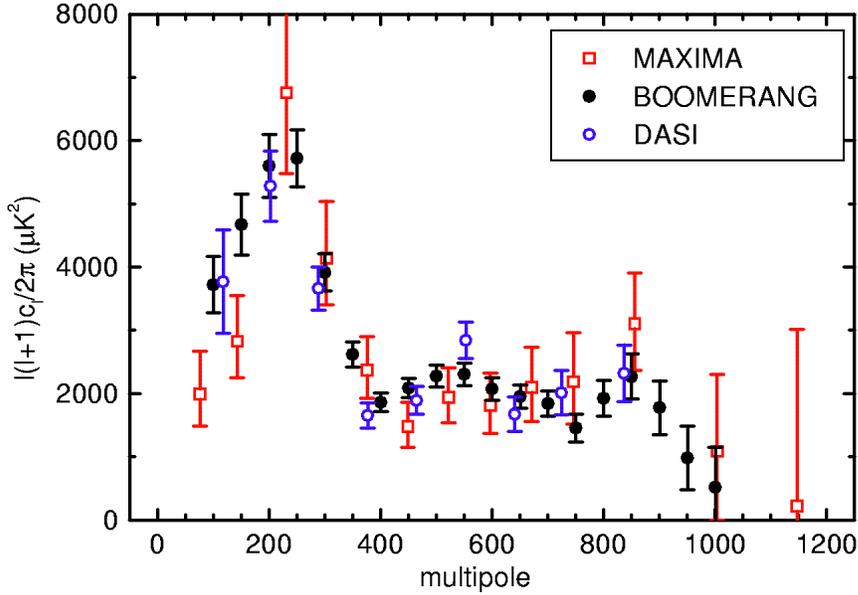


Figure 2: CMB anisotropy power spectrum detected by BOOMERanG, MAXIMA and DASI. Approximately uncorrelated bandpowers are plotted for each of the experiments. The error bars represent statistical errors only. A peak at $\ell \sim 200$ is evident in all datasets. This is a strong indication of a critical density of the Universe.

in the sky coverage of the BOOMERanG and DASI data should be taken into account for a more quantitative comparison. Such a comparison will be the best argument to exclude significant systematic effects in the three spectra.

7 Measuring Ω from the power spectrum

ℓ_{peak} has been evaluated from the BOOMERanG data by means of a quadratic fit to the power spectrum. We find $\ell_1 = (213^{+10}_{-13})$ ³². In the framework of inflationary adiabatic cold dark matter models this location of the peak strongly suggests a flat geometry of the Universe. More accurate statements require a through analysis of the full power spectrum dataset. A bayesian likelihood

analysis has been carried out in order to constrain instrumental and cosmological parameters given the measured power spectrum, the COBE power spectrum data at low multipoles, and a set of prior distributions for the parameters. A database of power spectra, for several millions of cases with different combinations of the values of the cosmological parameters, has been computed. The ranges and sampling selected for each parameter are wide enough to cover in detail the relevant parameters space. For each model the likelihood of the data, given the model and the assumed prior distributions, has been computed. Uniform prior distributions for all the assumed parameters have been used. The results obtained from the data of ³⁰⁾ are robust with respect to changes in the priors. Using only the prior $0.45 < h < 0.9$, the Bayesian determination of Ω is $\Omega = 1.02^{+0.05}_{-0.06}$. This result comes from a marginalization over all the other parameters and over the uncertainties in the gain calibration and in the beam FWHM. This determination of Ω definitely points towards a Euclidean geometry of our Universe. The BOOMERanG result above is in very good agreement with the independent determinations of DASI ($\Omega = 1.04 \pm 0.06$) ³³⁾ and MAXIMA ($\Omega = 0.90 \pm 0.15$) ⁴³⁾.

The extension of the multipoles coverage up to $\ell \sim 1000$ allows breaking the important degeneracy between $\Omega_b h^2$ and n . From BOOMERanG $\Omega_b h^2 = 0.022^{+0.004}_{-0.003}$ and $n_s = 0.96^{+0.10}_{-0.09}$. The quality of the data is so good that just the addition of the weak prior on h allows breaking the degeneracy between Ω_Λ and Ω_M . We find $\Omega_\Lambda = 0.51^{+0.23}_{-0.20}$ and $\Omega_{DM} h^2 = 0.13 \pm 0.05$. The addition of either a prior on the large scale structure (through the parameters σ_8 and Γ) or a prior on the SN1a observations, or even just a more restrictive prior on $h = 0.71 \pm 0.08$ produce very significant detections of Ω_Λ : $\Omega_\Lambda = (0.55 \pm 0.09)$, $(0.73^{+0.07}_{-0.10})$, $(0.62^{+0.10}_{-0.18})$ respectively ³⁰⁾, ³²⁾.

8 Conclusions

The CMB anisotropy experiments have recently resolved horizon and sub-horizon structures at recombination. Their typical size is $\sim 1^\circ$, strongly suggesting a Universe with Euclidean geometry and critical density. The density of the Universe has significant contributions from dark energy (about 70%), dark matter (about 25%) and baryons (less than 4%). Independent cosmological observations point to this unexpected composition of the Universe. Understanding the nature of dark matter and dark energy is now the most important

goal of cosmological research.

9 Acknowledgements

This work has been supported by ASI, PNRA and University of Rome La Sapienza. The BOOMERanG program has also been supported in USA by NSF and NASA, in UK by PPARC, and by CIAR and NSERC in Canada. We thank the organizers of the ISSS for wonderful hospitality.

References

1. J.E. Peebles *et al.*, Principles of Physical Cosmology, Princeton University Press, (1993).
2. J.A. Peacock, Cosmological Physics, Cambridge University Press, (1999).
3. W. Freedman, in "David Shramm memorial volume", Physics Reports, astro-ph/9909076, (2000).
4. E.W. Kolb and Turner M.S, The Early Universe, Addison-Wesley, (1990).
5. Fukugita *et al.*, Ap.J. **503**, 518 (1998).
6. Bahcall *et al.*, astro-ph/9906463 (1999).
7. Wittman *et al.*, Nature **405**, 143 (2000).
8. R. Juszkiewicz *et al.*, Science **287**, 109 (2000).
9. <http://opposite.stsci.edu/pubinfo/pr/2000/08/pr-photos.html>
10. A.G. Riess *et al.* Ap.J. **116**, 1009 (1998).
11. S. Perlmutter *et al.*, Nature **391**, 51 (1998).
12. L.M. Krauss, M.S. Turner, Gen.Rel.Grav. **31**, 1453 (1999).
13. Weinberg S., Gravitation and Cosmology, (1981).
14. Peebles P.J.E, Principles of Physical Cosmology, Princeton Series in Physics, (1994).
15. Sachs R.K., Wolfe, A.M., Ap.J. **147**, 73 (1967).

16. Mather J. *et al.*, Ap.J. **354**, L37 (1990).
17. Bond D., and Efstathiou G., astro-ph/9807103, (1998).
18. Weinberg S., astro-ph/0006276, (2000).
19. Melchiorri A., and Griffith, X, astro-ph/0011147, (2000).
20. Hu W., and Dodelson A., Ann.Rev.Astron.Astrophys., astro-ph/0110414, (2002).
21. <http://map.gsfc.nasa.gov>
22. <http://astro.estec.esa.nl/Planck>
23. G. Smoot *et al.*, Ap.J. **396**, L1 (1992).
24. C. Bennett *et al.*, Ap.J. **464**, L1 (1996).
25. A. Miller *et al.*, Ap.J. **524**, L1 (1999).
26. E. Torbet *et al.*, Ap.J. **521**, L79 (1999).
27. P. Matuszewska-Janusz *et al.*, Ap.J. **536**, L59 (2000).
28. P. de Bernardis *et al.*, Nature **404**, 955-959 (2000).
29. S. Hanany *et al.* Ap.J. **545**, L5 (2000).
30. B. Netterfield *et al.* submitted to Ap.J., astro-ph/0104460, (2001).
31. E.M. Leitch *et al.* astro-ph/0104488 (2001).
32. P. de Bernardis *et al.*, Ap.J. in press, astro-ph/0105296 (2002).
33. C. Pryke *et al.*, astro-ph/0104490 (2002).
34. A.T. Lee *et al.*, astro-ph/0104459 (2001).
35. S. Masi *et al.*, in "3K cosmology", AIP Conf. Proc. 476, 237 (1999); astro-ph/9911520
36. B.P. Crill Ph.D. thesis, Caltech. (2000).
37. F. Piacentini *et al.* astro-ph/0105148

38. D.J. Schlegel *et al.*, *Ap.J.* **500**, 525 (1998).
39. P. Mauskopf *et al.*, *Appl.Opt.* **36**, 765 (1997).
40. S. Masi *et al.*, *Cryogenics* **38**, 319 (1998).
41. S. Masi *et al.*, *Cryogenics* **39**, 217 (1999).
42. S. Masi *et al.*, *Ap.J.* **553**, L93 (2001).
43. A. Balbi *et al.* astro-ph/0005124, (2001).

THE NATURE OF DARK MATTER

Joel R. Primack

Physics Department, University of California, Santa Cruz, CA 95064 USA

ABSTRACT

Cold Dark Matter (CDM) has become the standard modern theory of cosmological structure formation. Its predictions appear to be in good agreement with data on large scales, and it naturally accounts for many properties of galaxies. But despite its many successes, there has been concern about CDM on small scales because of the possible contradiction between the linearly rising rotation curves observed in some dark-matter-dominated galaxies vs. the $1/r$ density cusps at the centers of simulated CDM halos. Other CDM issues on small scales include the very large number of small satellite halos in simulations, far more than the number of small galaxies observed locally, and problems concerning the angular momentum of the baryons in dark matter halos. The latest data and simulations have lessened, although not entirely resolved, these concerns. Meanwhile, the main alternatives to CDM that have been considered to solve these problems, self-interacting dark matter (SIDM) and warm dark matter (WDM), have been found to have serious drawbacks.¹

¹This paper is an significantly updated revision of ¹).

1 Introduction

Sometimes a theory is proposed in relatively early stages of the development of a scientific field, and this theory turns out to be not only a useful paradigm for the further development of the field — it also survives confrontation with a vast amount of data, and becomes accepted as the standard theory. This happened with General Relativity ²⁾, and it seems to be happening now with general relativistic cosmology. It appears that the universe on the largest scales can indeed be described by three numbers:

- $H_0 \equiv 100h\text{kms}^{-1}\text{Mpc}^{-1}$, the Hubble parameter (expansion rate of the universe) at the present epoch,
- $\Omega_m \equiv \rho/\rho_c$, the density of matter ρ in units of critical density $\rho_c \equiv 3H_0^2(8\pi G)^{-1} = 2.78 \times 10^{11}h^2M_\odot\text{Mpc}^{-3}$, and
- $\Omega_\Lambda \equiv \Lambda(3H_0^2)^{-1}$, the corresponding quantity for the cosmological constant.

The currently measured values of these and other key parameters are summarized in the Table below. It remains to be seen whether the “dark energy” represented by the cosmological constant Λ is really constant, or is perhaps instead a consequence of the dynamics of some fundamental field as in “quintessence” theories ³⁾.

In particle physics, the first unified theory of the weak and electromagnetic interactions ⁴⁾ had as its fundamental bosons just the carriers of the charged weak interactions W^+ , W^- , and the photon γ . The next such theory ⁵⁾ had a slightly more complicated pattern of gauge bosons — a triplet plus a singlet, out of which came not only W^+ , W^- , and γ , but also the neutral weak boson Z^0 , and correspondingly an extra free parameter, the “Weinberg angle.” It was of course this latter $SU(2)\times U(1)$ theory which has now become part of the Standard Model of particle physics. During the early 1970s, however, when the experimental data were just becoming available and some of the data appeared to contradict the $SU(2)\times U(1)$ theory, many other more complicated theories were proposed, even by Weinberg ⁶⁾, but all these more complicated theories ultimately fell by the wayside.

The development of theories of dark matter may follow a similar pattern. By the late 1970s it was becoming clear both that a great deal of dark matter exists ⁷⁾ and that the cosmic microwave background (CMB) fluctuation amplitude is smaller than that predicted in a baryonic universe. The first non-baryonic dark matter candidate to be investigated in detail was light neutrinos — what we now call “hot dark matter” (HDM). This dark matter is called “hot” because at one year after the big bang, when the horizon first encompassed the amount of matter in a large galaxy like our own (about $10^{12} M_{\odot}$) and the temperature was about 1 keV ⁸⁾, neutrinos with masses in the eV range would have been highly relativistic.

It is hardly surprising that HDM was worked out first. Neutrinos were known to exist, after all, and an experiment in Moscow that had measured a mass for the electron neutrino $m(\nu_e) \approx 20$ eV (corresponding to $\Omega_m \approx 1$ if h were as small as ~ 0.5 , since $\Omega_{\nu} = m(\nu_e)(92h^2\text{eV})^{-1}$) had motivated especially Zel’dovich and his colleagues to work out the implications of HDM with a Zel’dovich spectrum ($P_p(k) = Ak^n$ with $n = 1$) of adiabatic primordial fluctuations. But improved experiments subsequently have only produced upper limits for $m(\nu_e)$, currently about 3 eV ⁹⁾, and the predictions of the adiabatic HDM model are clearly inconsistent with the observed universe ^{10, 11)}.

Cold Dark Matter (CDM) was worked out as the problems with HDM were beginning to become clear. CDM assumes that the dark matter is mostly cold — i.e., with negligible thermal velocities in the early universe, either because the dark matter particles are weakly interacting massive particles (WIMPs) with mass $\sim 10^2$ GeV, or alternatively because they are produced without a thermal distribution of velocities, as is the case with axions. The CDM theory also assumes, like HDM, that the fluctuations in the dark matter have a nearly Zel’dovich spectrum of adiabatic fluctuations. Considering that the CDM model of structure formation in the universe was proposed almost twenty years ago ^{12, 13, 14)}, its successes are nothing short of amazing. As I will discuss, the Λ CDM variant of CDM with $\Omega_m = 1 - \Omega_{\Lambda} \approx 0.3$ appears to be in good agreement with the available data on large scales. Issues that have arisen on smaller scales, such as the centers of dark matter halos and the numbers of small satellites, have prompted people to propose a wide variety of alternatives to CDM, such as self-interacting dark matter (SIDM) ¹⁵⁾. It remains to be seen whether such alternative theories with extra parameters ac-

tually turn out to be in better agreement with data. As I will discuss below, it now appears that SIDM is probably ruled out, while the small-scale predictions of CDM may be in better agreement with the latest data than appeared to be the case as recently as a year ago.

In the next section I will briefly review the current observations and the successes of Λ CDM on large scales, and then I will discuss the possible problems on small scales.

2 Cosmological Parameters and Observations on Large Scales

The table below² summarizes the current observational information about the cosmological parameters, with estimated 1σ errors. The quantities in brackets have been deduced using at least some of the Λ CDM assumptions. It is apparent that there is impressive agreement between the values of the parameters determined by various methods, including those based on Λ CDM. In particular, (A) several different approaches (some of which are discussed further below) all suggest that $\Omega_m \approx 0.3$; (B) the location of the first acoustic peak in the CMB angular anisotropy power spectrum, now very well determined independently by the BOOMERANG¹⁷⁾ and MAXIMA1¹⁸⁾ balloon data^{19, 20)} and by the DASI interferometer at the South Pole²¹⁾, implies that $\Omega_m + \Omega_\Lambda \approx 1$; and (C) the data on supernovae of Type Ia (SNIa) at redshifts $z = 0.4 - 1.2$ from two independent groups imply that $\Omega_\Lambda - \frac{4}{3}\Omega \approx \frac{1}{3}$. Any two of these three results then imply that $\Omega_\Lambda \approx 0.7$. The 1σ errors in these determinations are about 0.1.

Questions have been raised about the reliability of the high-redshift SNIa results, especially the possibilities that the SNIa properties at high redshift might not be sufficiently similar to those nearby to use them as standard candles, and that there might be “grey” dust (which would make the SNIa dimmer but not change their colors). Although the available evidence disfavors these possibilities,³ additional observations are needed on SNIa at high redshift, both

²Further discussion and references are given in¹⁶⁾.

³For example, SNIa at $z = 1.2$ and ~ 1.7 apparently have the brightness expected in a Λ CDM cosmology but are brighter than would be expected with grey dust, and the infrared brightness of a nearer SNIa is also inconsistent with grey dust²²⁾.

Table 1: Cosmological Parameters [results assuming Λ CDM in brackets]

H_0	$= 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$, $h = 0.7 \pm 0.08$
t_0	$= 13 \pm 2 \text{ Gyr}$ (from globular clusters) $= [12 \pm 2 \text{ Gyr}$ from expansion age, Λ CDM model]
Ω_b	$= (0.039 \pm 0.006) h_{70}^{-2}$ (from D/H) $> [0.035 h_{70}^{-2}$ from Ly α forest opacity]
Ω_m	$= 0.4 \pm 0.2$ (from cluster baryons) $= [0.34 \pm 0.1$ from Ly α forest $P(k)$] $= [0.4 \pm 0.2$ from cluster evolution] $\approx \frac{3}{4} \Omega_\Lambda - \frac{1}{4} \pm \frac{1}{8}$ from SN Ia
$\Omega_m + \Omega_\Lambda$	$= 1.02 \pm 0.06$ (from CMB peak location)
Ω_Λ	$= 0.73 \pm 0.08$ (from previous two lines) < 0.73 (2σ) from radio QSO lensing
Ω_ν	$\gtrsim 0.001$ (from SuperKamiokande data) $\lesssim [0.1$ in Λ CDM-type models]

to control systematic effects and to see whether the dark energy is just a cosmological constant or is perhaps instead changing with redshift as expected in “quintessence” models ³⁾. Such data could be obtained by the proposed SuperNova Acceleration Probe (SNAP) satellite ²³⁾, whose Gigapixel optical camera and other instruments would also produce much other useful data. But it is important to appreciate that, independently of (C) SNIa, (A) cluster and other evidence for $\Omega_m \approx 0.3$, together with (B) $\sim 1^\circ$ CMB evidence for $\Omega_m + \Omega_\Lambda \approx 1$, imply that $\Omega_\Lambda \approx 0.7$.

All methods for determining the Hubble parameter now give compatible results, confirming our confidence that this crucial parameter has now been measured robustly to a 1σ accuracy of about 10%. The final result ²⁴⁾ from the Hubble Key Project on the Extragalactic Distance Scale is $72 \pm 8 \text{ km s}^{-1} \text{ Mpc}^{-1}$, or $h = 0.72 \pm 0.08$, where the stated error is dominated by one systematic uncertainty, the distance to the Large Magellanic Cloud (used to calibrate the Cepheid period-luminosity relationship). The most accurate of the direct methods for measuring distances d to distant objects, giving the Hubble parameter directly as $H_0 = d/v$ where the velocity is determined by the redshift, are (1) time delays between luminosity variations in different gravitationally lensed images of distant quasars, giving $h \approx 0.65$, and (2) the Sunyaev-Zel’dovich

effect (Compton scattering of the CMB by the hot electrons in clusters of galaxies), giving $h \approx 0.63$ ^{25, 24}). For the rest of this article, I will take $h = 0.7$ whenever I need to use an explicit value, and express results in terms of $h_{70} \equiv H_0/70\text{km s}^{-1} \text{Mpc}^{-1}$.

For a Λ CDM universe with $\Omega_m = (0.2)0.3(0.4, 0.5)$, the expansion age is $t_0 = (15.0)13.47(12.41, 11.61)h_{70}^{-1}$ Gyr. Thus for $\Omega_m \approx 0.3 - 0.4$ and $h \approx 0.7$, there is excellent agreement with the latest estimates of the ages of the oldest globular cluster stars in the Milky Way, both from their Main Sequence turnoff luminosities ²⁶), giving $12 - 13 \pm 2$ Gyr, and using the thorium and uranium radioactive decay chronometers ²⁷), giving 14 ± 3 Gyr and 12.5 ± 3 Gyr, respectively.

The simplest and clearest argument that $\Omega_m \approx 1/3$ comes from comparing the baryon abundance in clusters $f_b \equiv M_b/M_{tot}$ to that in the universe as a whole Ω_b/Ω_m , as emphasized by White et al. ²⁸). Since clusters are evidently formed from the gravitational collapse of a region of radius ~ 10 Mpc, they should represent a fair sample of both baryons and dark matter. This is confirmed in CDM simulations ²⁹). The fair sample hypothesis implies that

$$\Omega_m = \frac{\Omega_b}{f_b} = 0.3 \left(\frac{\Omega_b}{0.04} \right) \left(\frac{0.13}{f_b} \right). \quad (1)$$

We can use this to determine Ω_m using the baryon abundance $\Omega_b h^2 = 0.019 \pm 0.0024$ (95% C.L.) from the measurement of the deuterium abundance in high-redshift Lyman limit systems ^{30, 31}). Using X-ray data from an X-ray flux limited sample of clusters to estimate the baryon fraction $f_b = 0.075h^{-3/2}$ gives ³²) $\Omega_m = 0.25h^{-1/2} = 0.3 \pm 0.1$ (using $h = 0.70 \pm 0.08$). Estimating the baryon fraction using Sunyaev-Zel'dovich measurements of a sample of 18 clusters gives $f_b = 0.077h^{-1}$ ²⁵), and implies $\Omega_m = 0.25h^{-1} = 0.36 \pm 0.1$.

There is another way to use clusters to measure Ω_m , which takes advantage of the fact that the redshift at which structures form depends strongly on Ω_m . This happens because in a low-density universe the growth rate of fluctuations slows when, on the right hand side of the Friedmann equation,

$$H^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3}, \quad (2)$$

the first (matter) term becomes smaller than either the second (curvature) term (for the case of an open universe) or the third (cosmological constant) term.

As I have already discussed, the Λ term appears to be dominant now; note that if we evaluate the Friedmann equation at the present epoch and divide both sides by H_0^2 , the resulting equation is just

$$1 = \Omega_m + \Omega_k + \Omega_\Lambda . \quad (3)$$

Therefore, if we normalize the fluctuation power spectrum $P(k)$ for an $\Omega_m = 1$ (Einstein-de Sitter) cosmology and for a Λ CDM one by choosing σ_8 so that each is consistent with COBE and has the same abundance of clusters today, then at higher redshifts the low- Ω_m universe will have a higher comoving number density of clusters. Probably the most reliable way of comparing clusters nearby with those at higher redshift uses the cluster X-ray temperatures; the latest results, comparing 14 clusters at an average redshift of 0.38 with 25 nearby clusters, give $\Omega_m = 0.44 \pm 0.12$ (33, 34). There is greater leverage in this test if one can use higher redshift clusters, but the challenge is to find large samples with well understood cluster selection and properties. The largest such sample now available is from the Las Comanas Distant Cluster Survey, which goes to redshifts ~ 1 , from which the preliminary result is $\Omega_m = 0.30 \pm 0.12$ (90% CL) (35).

3 Further Successes of Λ CDM

We have already seen that Λ CDM correctly predicts the abundances of clusters nearby and at $z \lesssim 1$ within the current uncertainties in the values of the parameters. It is even consistent with $P(k)$ from the Ly α forest (36) and from CMB anisotropies. Low- Ω_m CDM predicts that the amplitude of the power spectrum $P(k)$ is rather large for $k \lesssim 0.02h/\text{Mpc}^{-1}$, i.e. on size scales larger (k smaller) than the peak in $P(k)$. The largest-scale surveys, 2dF and SDSS, should be able to measure $P(k)$ on these scales and test this crucial prediction soon; preliminary results are encouraging (37).

The hierarchical structure formation which is inherent in CDM already explains why most stars are in big galaxies like the Milky Way (14): smaller galaxies merge to form these larger ones, but the gas in still larger structures takes too long to cool to form still larger galaxies, so these larger structures — the largest bound systems in the universe — become groups and clusters instead of galaxies.

What about the more detailed predictions of Λ CDM, for example on the spatial distribution of galaxies. On large scales, there appears to be a pretty good match. In order to investigate such questions quantitatively on the smaller scales where the best data is available it is essential to do N-body simulations, since the mass fluctuations $\delta\rho/\rho$ are nonlinear on the few-Mpc scales that are relevant. My colleagues and I were initially concerned that Λ CDM would fail this test, ³⁸⁾ since the dark matter power spectrum $P_{dm}(k)$ in Λ CDM, and its Fourier transform the correlation function $\xi_{dm}(r)$, are seriously in disagreement with the galaxy data $P_g(k)$ and $\xi_g(r)$. One way of describing this is to say that scale-dependent antibiasing is required for Λ CDM to agree with observations. That is, the bias parameter $b(r) \equiv [\xi_g(r)/\xi_{dm}(r)]^{1/2}$, which is about unity on large scales, must decrease to less than 1/2 on scales of a few Mpc ^{38, 39)}. This was the opposite of what was expected: galaxies were generally thought to be more correlated than the dark matter on small scales. However, when it became possible to do simulations of sufficiently high resolution to identify the dark matter halos that would host visible galaxies ^{40, 41)}, it turned out that their correlation function is essentially identical with that of observed galaxies! This is illustrated in Fig. 1.

Jim Peebles, who largely initiated the study of galaxy correlations and first showed that $\xi_g(r) \approx (r/r_0)^{-1.8}$ with $r_0 \approx 5h^{-1}\text{Mpc}$ ⁴⁸⁾, thought that this simple power law must be telling us something fundamental about cosmology. However, it now appears that the power law ξ_g arises because of a coincidence – an interplay between the non-power-law $\xi_{dm}(r)$ (see Fig. 1) and the decreasing survival probability of dark matter halos in dense regions because of their destruction and merging. But the essential lesson is that Λ CDM correctly predicts the observed $\xi_g(r)$.

The same theory also predicts the number density of galaxies. Using the observed correlations between galaxy luminosity and internal velocity, known as the Tully-Fisher and Faber-Jackson relations for spiral and elliptical galaxies respectively, it is possible to convert observed galaxy luminosity functions into approximate galaxy velocity functions, which describe the number of galaxies per unit volume as a function of their internal velocity. The velocity function of dark matter halos is robustly predicted by N-body simulations for CDM-type theories, but to connect it with the observed internal velocities of bright galaxies it is necessary to correct for the infall of the baryons in these galaxies

43, 44, 45), which must have happened to create their bright centers and disks. When we did this it appeared that Λ CDM with $\Omega_m = 0.3$ predicts perhaps too many dark halos compared with the number of observed galaxies with internal rotation velocities $V \approx 200 \text{ km s}^{-1}$ (46, 47). While the latest results from the

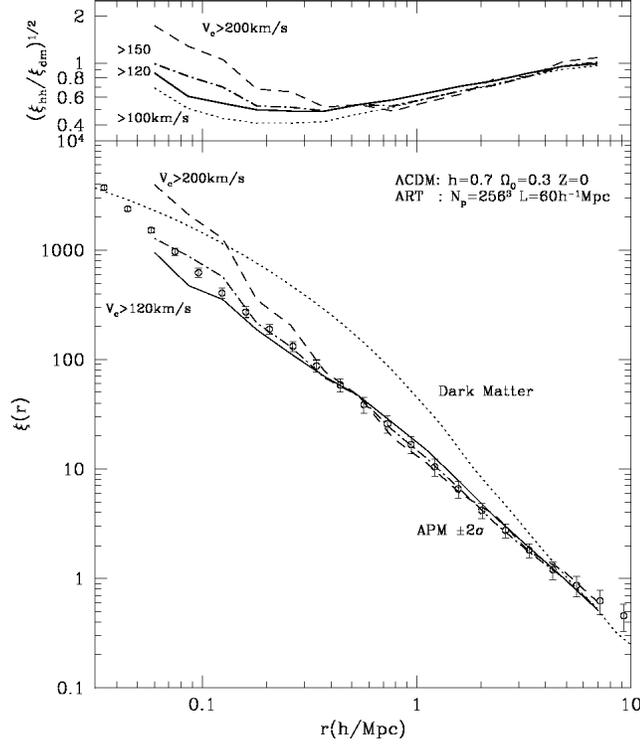


Figure 1: *Bottom panel:* Comparison of the halo correlation function in an Λ CDM simulation with the correlation function of the APM galaxies (42). The *dotted* curve shows the dark matter correlation function. Results for halos with maximum circular velocity larger than 120 km s^{-1} , 150 km s^{-1} , and 200 km s^{-1} are presented by the *solid*, *dot-dashed*, and *dashed* curves, respectively. Note that at scales $\geq 0.3 h^{-1} \text{ Mpc}$ the halo correlation function does not depend on the limit in the maximum circular velocity. *Top panel:* Dependence of bias on scale and maximum circular velocity. The curve labeling is the same as in the bottom panel, except that the *dotted* curve now represents the bias of halos with $V_{max} > 100 \text{ km s}^{-1}$. From Colin et al. (41).

big surveys now underway appear to be in better agreement with these Λ CDM predictions (49, 50), this is an important issue that is being investigated in detail (51).

The problem just mentioned of accounting for baryonic infall is just one example of the hydrodynamical phenomena that must be taken into account in order to make realistic predictions of galaxy properties in cosmological theories. Unfortunately, the crucial processes of especially star formation and supernova feedback are not yet well enough understood to allow reliable calculations. Therefore, rather than trying to understand galaxy formation from full-scale hydrodynamic simulations (for example (52)), more progress has been made via the simpler approach of semi-analytic modelling of galaxy formation (initiated by White and Frenk (53, 54, 55), recently reviewed and extended by Rachel Somerville and me (56)). The computational efficiency of SAMs permits detailed exploration of the effects of the cosmological parameters, as well as the parameters that control star formation and supernova feedback. We have shown (56) that both flat and open CDM-type models with $\Omega_m = 0.3-0.5$ predict galaxy luminosity functions and Tully-Fisher relations that are in good agreement with observations. Including the effects of (proto-)galaxy interactions at high redshift in SAMs allows us to account for the observed properties of high-redshift galaxies, but only for $\Omega_m \approx 0.3-0.5$ (57). Models with $\Omega_m = 1$ and realistic power spectra produce far too few galaxies at high redshift, essentially because of the fluctuation growth rate argument mentioned above.

In order to tell whether Λ CDM accounts in detail for galaxy properties, it is essential to model the dark halos accurately. The Navarro-Frenk-White (NFW) (58) density profile $\rho_{NFW}(r) \propto r^{-1}(r+r_s)^{-2}$ is a good representation of typical dark matter halos of galactic mass, except possibly in their very centers (§4). Comparing simulations of the same halo with numbers of particles ranging from $\sim 10^3$ to $\sim 10^6$, my colleagues and I have also shown (59) that r_s , the radius where the log-slope is -2, can be determined accurately for halos with as few as $\sim 10^3$ particles. Based on a study of thousands of halos at many redshifts in an Adaptive Refinement Tree (ART) (60) simulation of the Λ CDM cosmology, we (61) found that the concentration $c_{\text{vir}} \equiv R_{\text{vir}}/r_s$ has a log-normal distribution, with $1\sigma \Delta(\log c_{\text{vir}}) = 0.14$ at a given mass (62, 63). This scatter in concentration results in a scatter in maximum rotation velocities of $\Delta V_{\text{max}}/V_{\text{max}} = 0.12$; thus the distribution of halo concentrations has as

large an effect on galaxy rotation curves shapes as the well-known log-normal distribution of halo spin parameters λ . Frank van den Bosch (64) showed, based on a semi-analytic model for galaxy formation including the NFW profile and supernova feedback, that the spread in λ mainly results in movement along the Tully-Fisher line, while the spread in concentration results in dispersion perpendicular to the Tully-Fisher relation. Remarkably, he found that the dispersion in Λ CDM halo concentrations produces a Tully-Fisher dispersion that is consistent with the observed one.⁴

4 Halo Centers

Already in the early 1990s, high resolution simulations of individual galaxy halos in CDM were finding $\rho(r) \sim r^{-\alpha}$ with $\alpha \sim 1$. This behavior implies that the rotation velocity at the centers of galaxies should increase as $r^{1/2}$, but the data, especially that on dark-matter-dominated dwarf galaxies, instead showed a linear increase with radius, corresponding to roughly constant density in the centers of galaxies. This disagreement of theory with data led to concern that CDM might be in serious trouble (65, 66).

Subsequently, NFW (58) found that halos in all variants of CDM are well fit by the $\rho_{NFW}(r)$ given above, while Moore’s group proposed an alternative $\rho_M(r) \propto r^{-3/2}(r + r_M)^{-3/2}$ based on a small number of very-high-resolution simulations of individual halos (67, 68, 69). Klypin and collaborators (including me) initially claimed that typical CDM halos have shallow inner profiles with $\alpha \approx 0.2$ (72), but we subsequently realized that the convergence tests that we had performed on these simulations were inadequate. We now have simulated a small number of galaxy-size halos with very high resolution (59), and find that they range between ρ_{NFW} and ρ_M . Actually, these two analytic density profiles are almost indistinguishable unless galaxies are probed at scales

⁴Actually, this was the case with the dispersion in concentration $\Delta(\log c_{vir}) = 0.1$ found for relaxed halos by Jing (62), while we (61) found the larger dispersion mentioned above. However Risa Wechsler, in her dissertation research with me (63), found that the dispersion in the concentration at fixed mass of the halos that have not had a major merger since redshift $z = 2$ (and could thus host a spiral galaxy) is consistent with that found by Jing. We also found that the median and dispersion of halo concentration as a function of mass and redshift are explained by the spread in halo mass accretion histories.

smaller than about 1 kpc, which is difficult but sometimes possible.

Meanwhile, the observational situation is improving. The rotation curves of dark matter dominated low surface brightness (LSB) galaxies were measured with radio telescopes during the 1990s, and the rotation velocity was typically found to rise linearly at their centers (70, 71, 72). But a group led by van den Bosch (73) showed that in many cases the large beam size of the radio telescopes did not adequately resolve the inner parts of the rotation curves, and they concluded that after correcting for beam smearing the data are on the whole consistent with expectations from CDM. Similar conclusions were

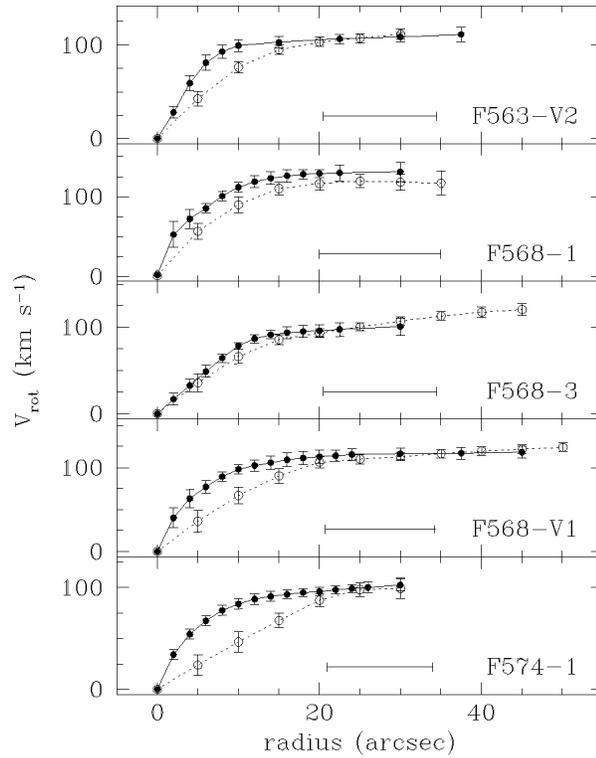


Figure 2: High resolution $H\alpha$ rotation curves (filled circles, solid lines) and HI rotation curves for the same galaxies (open circles, dotted lines) from Ref. 64. The horizontal bar shows the FWHM beam size of the HI observations. From Swaters, Madore, and Trehella (75).

reached for dwarf galaxies ⁷⁴). Swaters and collaborators showed that optical ($H\alpha$) rotation curves of some of the LSB galaxies rose significantly faster than the radio (HI) data on these same galaxies ⁷⁵) (see Fig. 2), and these rotation curves (except for F568-3) appear to be more consistent with NFW ⁷⁶). At a conference in March 2000 at the Institute for Theoretical Physics in Santa Barbara, Swaters also showed a $H\alpha$ rotation curve for the nearby dwarf galaxy DDO154, which had long been considered to be a problem for CDM ^{65, 66}); but the new, higher-resolution data appeared consistent with an inner density profile $\alpha \approx 1$.⁵

Very recently, a large set of high-resolution optical rotation curves has been analyzed for LSB galaxies, including many new observations ⁷⁷). The first conclusion that I reach in looking at the density profiles presented is that the NFW profile often appears to be a good fit down to about 1 kpc. However, some of these galaxies appear to have shallower density profiles at smaller radii. Of the 48 cases presented (representing 47 galaxies, since two different data sets are shown for F568-3), in a quarter of the cases the data do not probe inside 1 kpc, and in many of the remaining cases the resolution is not really adequate for definite conclusions, or the interpretation is complicated by the fact that the galaxies are nearly edge-on. Of the dozen cases where the inner profile is adequately probed, about half appear to be roughly consistent with the cuspy NFW profile (with fit $\alpha \gtrsim 0.5$), while half are shallower. This is not necessarily inconsistent with CDM, since observational biases such as seeing and slight misalignment of the slit lead to shallower profiles ⁷⁸). Perhaps it is significant that the cases where the innermost data points have the smallest errors are cuspiest.

I think that this data set may be consistent with an inner density profile $\alpha \sim 1$ but probably not steeper, so it is definitely inconsistent with the claims of the Moore group that $\alpha \gtrsim 1.5$. But very recent work by Navarro and collaborators ⁷⁹) has shown that Moore's simulations did not have adequate resolution to support their claimed steep central cusp; the highest-resolution simulations appear to be consistent with NFW, or even shallower with $\alpha \approx 0.75$. Further simulations and observations, including measurement of CO rotation curves ⁸⁰), may help to clarify the nature of the dark matter.

⁵Swaters (private communication) and Hoffman have subsequently confirmed this with better data, which they are preparing for publication.

It is something of a scandal that, after all these years of simulating dark matter halos, we still do not have a quantitative — or even a qualitative — theory explaining their radial density profiles. In her dissertation research ⁶³⁾, Risa Wechsler found that the central density profile and the value of r_s are typically established during the early, rapidly merging phase of halo evolution, and that, during the usually slower mass accretion afterward, r_s changes little. The mass added on the halo periphery increases R_{vir} , and thus the concentration $c_{vir} \equiv R_{vir}/r_s$. Now we want to understand this analytically. Earlier attempts to model the result of sequences of mergers (e.g., ^{81, 82)}) led to density profiles that depend strongly on the power spectrum of initial fluctuations, in conflict with simulations (e.g. ⁸³⁾). Perhaps it will be possible to improve on the simple analytic model of mass loss due to tidal stripping during satellite inspiral that we presented in ¹⁰¹⁾. Avishai Dekel and his students have recently shown that including the tidal puffing up of the inspiralling satellite before tidal stripping can perhaps account for the origin of the cusp seen in dissipationless simulations, independent of the power spectrum. They argue that the profile must be steeper than $\alpha = 1$ as long as enough satellites make it into the halo inner regions, simply because for flatter profiles the tidal force causes dilation rather than stripping. The proper modeling of the puffing and stripping in the merger process of CDM halos may also provide a theoretical framework for understanding the observed flat cores as a result of gas processes; work on this by Ari Maller and Dekel is in progress. Reionization and feedback into the baryonic component of small satellites would make their cores puff up before merging. This could cause them to be torn apart before they penetrate into the halo centers, and thus allow $\alpha < 1$ cores.

Another possible explanation for flatter central density profiles involving the baryonic component in galaxies has recently been proposed ⁸⁴⁾, in which the baryons form a bar that transfers angular momentum into the inner parts of the halo. It is not clear, however, that this effect could be very important in dark matter dominated dwarf and LSB galaxies that have small or nonexistent bulge components.

It would be interesting to see whether CDM can give a consistent account of the distribution of matter near the centers of big galaxies, but this is not easy to test. One might think that big bright galaxies like the Milky Way could help to test the predicted CDM profile, but the centers of such galaxies

are dominated by ordinary matter (stars) rather than dark matter.⁶

5 Too Much Substructure?

Another concern is that there are more dark halos in CDM simulations with circular velocity $V_c \lesssim 30 \text{ km s}^{-1}$ than there are low- V_c galaxies in the Local Group (87, 88). A natural solution to this problem was proposed by Bullock et al. (89), who pointed out that gas will not be able to cool in $V_c \lesssim 30 \text{ km s}^{-1}$ dark matter halos that collapse after the epoch of reionization, which occurred perhaps at redshift $z_{\text{reion}} \approx 6$ (90). When this is taken into account, the predicted number of small satellite galaxies in the Local Group is in good agreement with observations (89, 91). It is important to develop and test this idea further, and this is being done by James Bullock and by Rachel Somerville and their collaborators; the results to date (e.g. 92, 93) look rather promising. Other groups (e.g. 94, 95, 96) now agree that astrophysical effects will keep most of the subhalos dark. As a result, theories such as warm dark matter (WDM), which solve the supposed problem of too many satellites by decreasing the amount of small scale power, may end up predicting too few satellites when reionization and other astrophysical effects are taken into account (97).

The fact that high-resolution CDM simulations of galaxy-mass halos are full of subhalos has also led to concerns that all this substructure could prevent the resulting astrophysical objects from looking like actual galaxies (88). In particular, it is known that interaction with massive satellites can thicken or damage the thin stellar disks that are characteristic of spiral galaxies, after the disks have formed by dissipative gas processes. However, detailed simulations (98, 99) have shown that simpler calculations (100) had overestimated the extent to which small satellites could damage galactic disks. Only interaction with large satellites like the Large Magellanic Cloud could do serious damage. But the number of LMC-size and larger satellites is in good agreement with the number of predicted halos (87), which suggests that preventing disk damage will not lead to a separate constraint on halo substructure.

⁶Navarro and Steinmetz had claimed that the Milky Way is inconsistent with the NFW profile (85), but they have now shown that Λ CDM simulations with a proper fluctuation spectrum are actually consistent with the data (86).

6 Angular Momentum Problems

As part of James Bullock’s dissertation research, we found that the distribution of specific angular momentum in dark matter halos has a universal profile ¹⁰¹). But if the baryons have the same angular momentum distribution as the dark matter, this implies that there is too much baryonic material with low angular momentum to form the observed rotationally supported exponential disks ^{101, 102}). It has long been assumed (e.g. ^{43, 44}) that the baryons and dark matter in a halo start with a similar distribution, based on the idea that angular momentum arising from large-scale tidal torques will be similar across the entire halo. But as my colleagues and I argued recently, a key implication of our new picture of angular momentum growth by merging ¹⁰³) is that the DM and baryons will get different angular momentum distributions. For example, the lower density gas will be stripped by pressure and tidal forces from infalling satellites, and in big mergers the gaseous disks will partly become tidal tails. Feedback is also likely to play an important role, and Maller and Dekel (in preparation) have shown using a simple model that this can account for data on the angular momentum distribution in low surface brightness galaxies ¹⁰⁴).

A related concern is that high-resolution hydrodynamical simulations of galaxy formation lead to disks that are much too small, evidently because formation of baryonic substructure leads to too much transfer of angular momentum and energy from the baryons to the dark matter ¹⁰⁵). But if gas cooling is inhibited in the early universe, more realistic disks form ¹⁰⁶), more so in Λ CDM than in $\Omega_m = 1$ CDM ¹⁰⁷). Hydrodynamical simulations also indicate that this disk angular momentum problem may be resolved if small scale power is suppressed because the dark matter is warm rather than cold ¹⁰⁸), which I discuss next.

7 Alternatives to Λ CDM?

Because of the concerns just mentioned that CDM may predict higher densities and more substructure on small scales than is observed, many people have proposed alternatives to CDM. Two of these ideas that have been studied in the greatest detail are self-interacting dark matter (SIDM) ¹⁵) and warm dark matter (WDM).

Cold dark matter assumes that the dark matter particles have only weak

interactions with each other and with other particles. SIDM assumes that the dark matter particles have strong elastic scattering cross sections, but negligible annihilation or dissipation. The hope was that SIDM might suppress the formation of the dense central regions of dark matter halos, although the large cross sections might also lead to high thermal conductivity which drains energy from halo centers and could lead to core collapse ¹⁰⁹⁾, and which also causes evaporation of galaxy halos in clusters, resulting in violation of the observed “fundamental plane” correlations ¹¹⁰⁾. But in any case, self-interaction cross sections large enough to have a significant effect on the centers of galaxy-mass halos will make the centers of galaxy clusters more spherical ^{111, 112)} and perhaps also less dense ^{113, 114)} than gravitational lensing observations ¹¹⁵⁾ indicate.

Warm dark matter arises in particle physics theories in which the dark matter particles have relatively high thermal velocities, for example because their mass is $\lesssim 1$ keV ¹¹⁶⁾, comparable to the temperature about a year after the Big Bang when the horizon first encompassed the amount of dark matter in a large galaxy. Such a velocity distribution can suppress the formation of structure on small scales. Indeed, this leads to constraints on how low the WDM particle mass can be. From the requirement that there is enough small-scale power in the linear power spectrum to reproduce the observed properties of the Ly α forest in quasar spectra, it follows that this mass must exceed about 0.75 keV ¹¹⁷⁾. The requirement that there be enough small halos to host early galaxies to produce the floor in metallicity observed in the Ly α forest systems, and early galaxies and quasars to reionize the universe, probably implies a stronger lower limit on the WDM mass of at least 1 keV ¹¹⁸⁾. Simulations ^{119, 120)} do show that there will be far fewer small satellite halos with Λ WDM than Λ CDM. However, as I have already mentioned, inclusion of the effects of reionization may make the observed numbers of satellite galaxies consistent with the predictions of Λ CDM ⁸⁹⁾, in which case Λ WDM may predict too few small satellite galaxies. Lensing can be used to look for these subhalos ^{121, 122)} and may already indicate that there are more of them than expected in Λ WDM ¹²³⁾. Thus it appears likely that WDM does not solve all the problems it was invoked to solve, and may create new problems. Moreover, even with an initial power spectrum truncated on small scales, simulations appear to indicate that dark matter halos nevertheless have density profiles

much like those in CDM (124, 68, 86) (although doubts have been expressed about the reliability of such simulations because of numerical relaxation (125)). But WDM does lead to lower concentration halos in better agreement with observed rotation velocity curves (126, 127).

One theoretical direction that does appear very much worth investigating is Λ CDM with a tilt $n \sim 0.9$ in the primordial power spectrum $P_p(k) \propto k^n$ (128). Such $t\Lambda$ CDM cosmology is favored by recent measurements of the power spectrum of the Ly α forest (36) and appears to be consistent with the latest CMB measurements and all other available data (129). Our simple analytic model (61) predicts that the concentration of halos in $t\Lambda$ CDM will be approximately half that in LCDM, which appears to be true in a trial simulation by A. Kravtsov. While this does not resolve the cusp problem, it is a step in the right direction which may lessen the conflict with galaxy rotation curves.

8 Outlook

The successes of the CDM paradigm are remarkable. Except possibly for the density profiles at the centers of dwarf and low surface brightness galaxies, the predictions of Λ CDM appear to be in good agreement with the available observations. The disagreements between predictions and data at galaxy centers appear to occur on smaller scales than was once thought, but as the data improve it is possible that the discrepancies on $\lesssim 1$ kpc scales may ultimately show that CDM cannot be the correct theory of structure formation. However, it appears to be better than any alternative theory that has so far been studied, even though these alternative theories have more adjustable parameters.

This article started by discussing the analogy between the effort to understand dark matter and structure formation in modern cosmology and the effort to understand particle physics in the 1960s and 1970s. In both cases, the result was a “standard model” which has guided further work and led to great progress in both theory and observation/experiment. But in both cases, the standard model is not an ultimate theory, and the search is on for a better theory. In the case of particle physics, there is a leading candidate: supersymmetry, and perhaps ultimately string or M theory. Here the analogy fails, because I am not aware of any theory that has all the virtues of CDM but which avoids its possible failure at the centers of galaxies. The quest for such a theory is a worthwhile goal. But for many purposes, including studies of the

formation and evolution of galaxies and their large scale distribution, the CDM standard model may still remain very useful. And maybe it is even true.

9 Acknowledgements

I learned a great deal about the topics discussed here from my collaborators James Bullock, Avishai Dekel, Ricardo Flores, Anatoly Klypin, Andrey Kravtsov, Ari Maller, Rachel Somerville, and Risa Wechsler, and I also thank Frank van den Bosch, Volker Springel, Rob Swaters, and Simon White for very helpful discussions about the centers of galaxies. This work was supported by grants from NASA and NSF at UCSC, and by a Humboldt Award which supported my visits in fall 2001 to the Max Planck Institutes for Physics in Munich and Astrophysics in Garching, where I thank Leo Stodolsky and Simon White for hospitality. I also thank Aldo Morselli for inviting me to give these lectures at the ISSS, and for his patience waiting for me to send him these notes.

References

1. J.R. Primack, in *COSMO-2000, Proc. 4th International Workshop on Particle Physics and the Early Universe* (Jeju Island, Korea, September 2000), eds. J.E. Kim et al. (World Scientific, Singapore, 2001), pp. 1-18.
2. C.M. Will *Theory and experiment in gravitational physics* (Cambridge University Press, Cambridge, 1993).
3. P.J. Steinhardt *Physica Scripta* **185**, 177 (2000) and references therein.
4. J. Schwinger *Ann. of Phys.* **2**, 407 (1957). Cf. H. Georgi and S.L. Glashow, *Phys. Rev. Lett.* **28**, 1494 (1972).
5. S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967). Precursors include S.A. Bludman *Nuovo Cim.* **9**, 433 (1958), and S.L. Glashow, *Nucl. Phys.* **10**, 107 (1959).
6. S. Weinberg, *Phys. Rev.* **D5**, 1962 (1972).
7. S.M. Faber and J. Gallagher, *Ann. Rev. Astron. Astrophys.* **17**, 135 (1979).
8. J.R. Primack, in *Proceedings, International School of Physics "Enrico Fermi" XCII* (1984), ed. N. Cabibbo (North-Holland, 1987), p. 137.

9. See the on-line Particle Data Book <http://pdg.lbl.gov/2001/lxxx.html> (2001).
10. J.R. Primack and M.A.K. Gross, in ed. D. O. Caldwell, *Current Aspects of Neutrino Physics* (Springer, Berlin, 2001), p. 287.
11. J.R. Primack (2001) *Beam Line*, **31** (3 - Fall 2001), 50.
12. P.J.E. Peebles *ApJ* **263**, L1 (1982).
13. J.R. Primack and G.R. Blumenthal in *Formation and Evolution of Galaxies and Large Structures in the Universe*, eds. J. Audouze and J. Tran Thanh Van (Reidel, Dordrecht, 1983), p. 163.
14. G.R. Blumenthal, S.M. Faber, J.R. Primack, and M.J. Rees, *Nature* **311**, 517 (1984).
15. D.N. Spergel and P.J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000).
16. J.R. Primack, in *Proc. 4th International Symposium on Sources and Detection of Dark Matter in the Universe* (DM 2000), ed. D. Cline (Springer, Berlin, 2001), p. 3.
17. A.E. Lange et al., *Phys. Rev.* **D63**, 042001 (2000); P. de Bernardis et al., *ApJ* submitted, astro-ph/0105296 (2001).
18. A. Balbi et al., *ApJ*, **545**, L1 (2000), erratum **558**, L145 (2001); R. Stompor et al., *ApJ*, **561**, L7 (2001).
19. A.H. Jaffe et al., astro-ph/0007333 (2000).
20. J.R. Bond et al., astro-ph/0011378, in *Proc. IAU Symposium 201*, eds. A. Lasenby and A. Wilkinson (Astronomical Society of the Pacific, San Francisco, 2002).
21. C. Pryke et al., *ApJ* submitted, astro-ph/0104490 (2001).
22. A.G. Riess et al., *ApJ*, **536**, 62 (2000); *ApJ*, **560**, 49 (2001).
23. The SNAP proposal can be downloaded from <http://snap.lbl.gov>.
24. W.L. Freedman et al., *ApJ* **553**, 47 (2001).

25. J.E. Carlstrom et al., *Physica Scripta T* **85**, 148 (2000); J.E. Carlstrom et al., astro-ph/0103480, to appear in *Constructing the Universe with Clusters of Galaxies*, eds. F. Durret and G. Gerbal.
26. E. Carretta, R. Gratton, G. Clementini, and F. Fusi Pecci, *ApJ* **533**, 215 (2000).
27. C. Sneden et al., *ApJ* **536**, L85 (2000); R. Cayrel et al., *Nature* **409**, 691 (2001).
28. S.D.M. White, G. Efstathiou, C.S. and Frenk, *MNRAS* **262**, 1023 (1993).
29. A.E. Evrard, C.A. Metzler, and J.F. Navarro, *ApJ* **469**, 494 (1996).
30. D. Kirkman, D. Tytler, S. Burles, D. Lubin, and J.M. O'Meara *ApJ* **529**, 655 (2000).
31. D. Tytler, J.M. O'Meara, N. Suzuki, and D. Lubin, *Physica Scripta T***85**, 12 (2000).
32. J.J. Mohr, B. Mathiesen, and A.E. Evrard, *ApJ* **517**, 627 (1999).
33. V.R. Eke, S. Cole, C.S. Frenk, and J.P. Henry, *MNRAS* **298**, 1145 (1998).
34. J.P. Henry, *ApJ* **534**, 565 (2000).
35. A.H. Gonzalez, D. Zaritsky, J.J. Dalcanton, and A.E. Nelson, in *Clustering at High Redshift*, ed. A. Mazure et al., ASP Conference Series, Vol. 200, 416 (2000); A.H. Gonzalez (2000) UCSC Ph.D. dissertation.
36. D.H. Weinberg et al., (1999) *ApJ* **522**, 563; but cf. R.A. Croft et al., astro-ph/0012324 (2001).
37. W.J. Percival et al., (2dF), *MNRAS*, **327**, 1297 (2001); A.S. Szalay et al., (SDSS), astro-ph/0107419 (2001); cf. (2MASS) B. Allgood, G.R. Blumenthal, and J.R. Primack, in Proc. Marseille 2001 Conf. *Where's the matter? Tracing dark and and bright matter with the new generation of large scale surveys*, eds R. Treyer & L. Tresse, astro-ph/0109403 (2001).
38. A.A. Klypin, J.R. Primack, and J. Holtzman, *ApJ* **466**, 13 (1996).
39. A. Jenkins et al., *ApJ* **499**, 20 (1998).

40. A.A. Klypin, S. Gottlöber, A.V. Kravtsov, and A.M. Khokhlov *ApJ* **516**, 530 (1999).
41. P. Colin, A.A. Klypin, A.V. Kravtsov, and A.M. Khokhlov, *ApJ* **523**, 32 (1999).
42. C.M. Baugh, *MNRAS* **280**, 267 (1996).
43. G.R. Blumenthal, S.M. Faber, R. Flores, and J.R. Primack, *ApJ* **301**, 27 (1986); R. Flores, J.R. Primack, G.R. Blumenthal, and S.M. Faber, *ApJ* **412**, 443 (1993).
44. H.-J. Mo, S. Mao, and S.D.M. White, *MNRAS* **295**, 319 (1998).
45. C.S. Kochanek and M. White, *ApJ*, **559**, 531 (2001).
46. A.H. Gonzalez, K.A. Williams, J.S. Bullock, T.S. Kolatt, and J.R. Primack, *ApJ* **528**, 145 (2000).
47. S. Cole, C. Lacey, C. Baugh, and C.S. Frenk, *MNRAS* **319**, 168 (2000).
48. M. Davis and P.J.E. Peebles, *ApJ* **267**, 465 (1983).
49. N. Cross, S. Driver, and W. Couch (2dF Collaboration), astro-ph/0012165, *MNRAS* in press (2001).
50. M.R. Blanton et al., (SDSS Collaboration), *AJ* **121**, 2358 (2001).
51. M.A. Pahre et al. (MALIGN survey), AAS Meeting 197, #96.01 (2000).
52. D.H. Weinberg, L. Hernquist, and N. Katz, *ApJ* submitted astro-ph/0005340 (2000).
53. S.D.M. White and C.S. Frenk, *ApJ* **379**, 52 (1991).
54. G. Kauffmann, S.D.M. White, and B. Guiderdoni, *MNRAS* **264**, 201 (1993).
55. S. Cole et al., *MNRAS* **271**, 781 (1994).
56. R.S. Somerville and J.R. Primack, *MNRAS*, **310**, 1087 (1999).
57. R. S. Somerville, J.R. Primack, and S.M. Faber, *MNRAS* **320**, 504 (2001).

58. J.F. Navarro, C.D. Frenk, C.S. and S.D.M. White, *ApJ* **462**, 563 (1996);
ApJ **490**, 493 (1997).
59. A.A. Klypin, A.V. Kravtsov, J.S. Bullock, and J.R. Primack, *ApJ* **554**, 903
(2001).
60. A.V. Kravtsov, A.A. Klypin, and A.M. Khokhlov, *ApJS* **111**, 73 (1997).
61. J.S. Bullock, T.S. Kolatt, Y. Sigad, R.S. Somerville, A.V. Kravtsov, A.A.
Klypin, J.R. Primack, and A. Dekel, *MNRAS* **321**, 559 (2001).
62. Y. Jing, *ApJ* **535**, 30 (2000).
63. R.H. Wechsler, J.S. Bullock, J.R. Primack, A.V. Kravtsov, and A. Dekel,
ApJ in press, astro-ph/0108151 (2001).
64. F.C. van den Bosch (2000) *ApJ* **530**, 177.
65. R.A. Flores and J.R. Primack, *ApJ*, **427**, L1 (1994).
66. B. Moore, *Nature* **370**, 620 (1994).
67. B. Moore et al., *ApJ* **499**, L5 (1998).
68. B. Moore et al., *MNRAS* **310**, 1147 (1999).
69. S. Ghigna et al., *ApJ* **544**, 616 (2000).
70. W.J.G. de Blok, S.S. McGaugh, and J.M. van der Hulst, *MNRAS* **283**, 18
(1996).
71. S.S. McGaugh and W.J.G. de Blok, *ApJ* **499**, 41 (1998) and references
cited therein.
72. A.A. Klypin, A.V. Kravtsov, J.S. Bullock, and J.R. Primack, *ApJ* **502**, 48
(1998) and references cited therein.
73. F.C. van den Bosch, B.E. Robertson, J.J. Dalcanton, and W.J.G. de Blok,
AJ, **119**, 1579 (2000).
74. F. van den Bosch and R.A. Swaters, *MNRAS* **325**, 1017 (2001).
75. R.A. Swaters, B.F. Madore, and M. Trewhella, *ApJ* **531**, L107 (2000).

76. R.A. Swaters, in *Galaxy Disks and Disk Galaxies*, ed. J. G. Funes and E.M. Corsini, ASP Conf. Series, Vol. 230, p. 545 (2001).
77. W.J.G. de Blok, S.S. McGaugh, A. Bosma, and V.C. Rubin, *ApJ*, 552, L23 (2001); S.S. McGaugh, V.C. Rubin, and W.J.G. de Blok, *AJ* **122**, 2381 (2001); W.J.G. de Blok, S.S. McGaugh, and V.C. Rubin, *AJ* **122**, 2396 (2001).
78. R. Swaters, in Proc. *The Mass of Galaxies at Low and High Redshift*, Venice Oct 2001, eds. R. Bender and A. Renzini (Springer-Verlag, 2002).
79. J.F. Navarro, in IAU Symposium 208, *Astrophysical SuperComputing using Particles*, eds. J. Makino and P.Hut, astro-ph/0110680 (2001).
80. A.D. Bolatto, J.D. Simon, A. Leroy, and L. Blitz, *ApJ* in press astro-ph/0108050 (2001).
81. D. Syer and S.D.M. White, *MNRAS* **293**, 377 (1998).
82. A. Nusser and R.K. Sheth, *MNRAS* **303**, 685 (1999).
83. A. Huss, B. Jain, and M. Steinmetz, *ApJ* **517**, 64 (1999).
84. M.D. Weinberg and N. Katz, astro-ph/0110632 (2001).
85. J.F. Navarro and M. Steinmetz, *ApJ* **528**, 607 (2000).
86. V.R. Eke, J.F. Navarro, and M. Steinmetz, *ApJ* **554**, 114 (2001).
87. A.A. Klypin, A.V. Kravtsov, O. Valenzuela, and F. Prada, *ApJ* **522**, 82 (1999).
88. B. Moore et al., *ApJ* 524, L19 (1999).
89. J.S. Bullock, A.V. Kravtsov, and D.H. Weinberg, *ApJ* **539**, 517 (2000).
90. See X. Fan et al., *AJ* submitted, astro-ph/0111184 (2001), and references therein.
91. B. Moore, astro-ph/0103100, in *20th Texas Symposium*, eds. J. C. Wheeler and H. Martel, in press (2001).
92. J.S. Bullock, A.V. Kravtsov, and D.H. Weinberg, *ApJ* **548**, 33 (2001).

93. R.S. Somerville, *astro-ph/0107507* (2001).
94. E. Scannapieco, R.J. Thacker, and M. Davis, *ApJ* **557**, 605 (2001).
95. W.A. Chiu, N.Y. Gnedin, and J.P. Ostriker, *astro-ph/0103359* (2001).
96. A.J. Benson et al., *astro-ph/0108218* (2001).
97. R. Barkana, Z. Haiman, and J.P. Ostriker, *ApJ* **558**, 482 (2001).
98. I.R. Walker, J.C. Mihos, and L. Hernquist, *ApJ* **460**, 121 (1996).
99. H. Velazquez and S.D.M. White, *MNRAS* **304**, 254 (1999).
100. G. Toth and J.P. Ostriker, *ApJ* **389**, 5 (1992).
101. J.S. Bullock, A. Dekel, T.S. Kolatt, J.R. Primack, and R.S. Somerville, *ApJ* **555**, 240 (2001).
102. F. van den Bosch, *MNRAS*, **327**, 1334 (2001).
103. M. Vitvitska, A.A. Klypin, A.V. Kravtsov, J.S. Bullock, R.H. Wechsler, and J.R. Primack, *ApJ* submitted, *astro-ph/0105349*; A.H. Maller, A. Dekel, and R.S. Somerville, *MNRAS* submitted, *astro-ph/0105168* (2001).
104. F. van den Bosch, A. Burkert, and R.A. Swaters, *MNRAS*, **326**, 1205 (2001).
105. J.F. Navarro and M. Steinmetz, *ApJ* **478**, 13 (1997); **513**, 555 (1999); **538**, 477 (2000).
106. M.L. Weil, V.R. Eke, and g. Efstathiou, *MNRAS* **300**, 773 (1998).
107. V.R. Eke, G. Efstathiou, and L. Wright, *MNRAS* **315**, L18 (2000).
108. J. Sommer-Larsen and A. Dolgov, *ApJ* **551**, 608 (2001).
109. A. Burkert, *astro-ph/0012178* (2000).
110. O.Y. Gnedin and J.P. Ostriker, *ApJ* **561**, 61 (2001).
111. J. Miralda-Escude, *astro-ph/0002050* (2000).

112. N. Yoshida, V. Springel, S.D.M. White, and G. Tormen, *ApJ* **544**, L87 (2000).
113. M. Meneghetti et al., *MNRAS* **325**, 435 (2001).
114. J.S.B. Wyithe, E.L. Turner, and D.N. Spergel, *ApJ* **555**, 504 (2001).
115. See e.g. J.S. Arabadjis, M.W. Bautz, and G.P. Garmire, astro-ph/0109141 (2001).
116. H. Pagels and J.R. Primack, *Phys. Rev. Lett.* **48**, 223 (1982); G.R. Blumenthal, H. Pagels, and J.R. Primack, *Nature* **299**, 37 (1982).
117. V.K. Narayanan, D.N. Spergel, R. Davé, and C.P. Ma, *ApJ* **543**, L103 (2000).
118. Z. Haiman, R. Barkana, and J.P. Ostriker, astro-ph/0103050 (2001).
119. R. Colin, V. Avila-Reese, and O. Valenzuela, *ApJ* **542**, 622 (2000).
120. P. Bode, J.P. Ostriker, and N. Turok, *ApJ*, **556**, 93 (2001).
121. R.B. Metcalf and P. Madau, *ApJ* in press, astro-ph/0108224 (2001).
122. M. Chiba, *ApJ* in press, astro-ph/0109499 (2001).
123. N. Dalal and C.H. Kochanek, astro-ph/0111456 (2001).
124. A. Huss, B. Jain, and M. Steinmetz, *ApJ* **517**, 64 (1999).
125. J.J. Dalcanton and C.J. Hogan, *ApJ* **561**, 35 (2001).
126. V. Avila-Rees, P. Colin, O. Valenzuela, E. D'Onghia, and C. Firmani, *ApJ* **559**, 516 (2001).
127. S.M.K. Alam, J.S. Bullock, and D.H. Weinberg, *ApJ* submitted, astro-ph/0109392 (2001).
128. J.S. Bullock, in Proc. Marseille 2001 Conf. *Where's the matter? Tracing dark and bright matter with the new generation of large scale surveys*, eds R. Treyer & L. Tresse, astro-ph/0111005 (2001).
129. X. Wang, M. Tegmark, and M. Zaldarriaga, astro-ph/0105091 (2001).

EXOTIC COSMIC RAYS FROM PAIR ANNIHILATIONS OF DARK MATTER WIMPS

Piero Ullio
SISSA, via Beirut 4, 34014 Trieste, Italy

ABSTRACT

We consider the hypothesis that dark matter is made of weakly interacting massive particles (WIMPs) and describe how their pair annihilation in the galactic halo generates exotic cosmic ray fluxes. Features for generic WIMP models are reviewed, pointing out cases in which clear signatures arise. Implications from available and upcoming measurements are discussed.

1 Motivations

The case for non-baryonic dark matter making up most of the matter in the Universe is getting stronger and stronger. Observational evidence spans from galactic scales to the largest scales currently possible to probe, see, e.g., ¹). The hypothesis that dark matter is made of weakly interacting massive particles (WIMPs) is particularly appealing. On one hand, WIMPs arise naturally

in theories beyond the standard model of particle physics. Supersymmetry, which seems a necessary ingredient to unify gravity with the other fundamental forces, provides the best motivated WIMP dark matter candidate: the lightest supersymmetric particle, most probably a neutralino - a linear combination of the supersymmetric partners of the photon, the Z boson and neutral scalar Higgs particles. On the other hand, if a stable massive particle (which we indicate here generically with χ) were present in the early Universe in thermal equilibrium with ordinary matter, we expect to find a relic population of χ s in the Universe nowadays. The relevant effects are the χ pair annihilation rate into lighter particles (which maintains the thermal equilibrium) compared to the Universe expansion rate: when these two effects become roughly equal, χ s fall out of equilibrium and their cosmological density freezes in. It is not hard to show ²⁾, as a first order approximation, that the relic abundance of such particles scales as:

$$\Omega_\chi h^2 \simeq \frac{3 \cdot 10^{-27} \text{cm}^{-3} \text{s}^{-1}}{\langle \sigma_A v \rangle}, \quad (1)$$

where $\langle \sigma_A v \rangle$ is the thermally averaged total annihilation cross section of χ s into lighter Standard Model particles. It follows that if we require $\Omega_\chi h^2$ to be in the range currently favored for cold dark matter, about $0.1 - 0.2$ ³⁾, the coupling of χ to ordinary matter has to be of weak interaction strength. Reversing this argument, we see that WIMPs naturally conjugate motivations from particle physics and from cosmology.

The scheme we have just sketched defines the case of “thermal” relic WIMPs, numerous other plausible scenarios have been discussed recently, see, e.g., ⁴⁾: The WIMP dark matter candidates which arise in these schemes are generically indicated as “non-thermal”, as they rely on an alternative mechanism to provide the cold dark matter we see in the Universe today. Some degree of fine-tuning is usually involved, but, at the same time, these dark matter candidates imply richer phenomenological implications, including very large pair annihilation cross sections, interesting from the point of view of indirect detection of dark matter through cosmic ray searches.

2 Cosmic rays from WIMP annihilations

According to the current theory of structure formation, the Λ CDM model, cold dark matter drives the formation of all the structures we see in the Universe

today, with dark matter halos forming first and in a hierarchical fashion, and then with baryons collapsing in halo potential wells and giving rise to luminous components. If WIMPs are indeed the relics accounting for dark matter, we expect them to populate all dark halos, and in particular the halo of our own Galaxy, with radial density profiles described by the results from numerical N-body simulations ^{5, 6)} (maybe with slight modifications due to the interplay with baryons, which are not included in the simulations yet) and with velocities of the order of halo circular velocities (about 200 km s^{-1} for a spiral of the size of the Milky Way). Although such densities and velocities are much smaller than in the early Universe, there is still a finite probability for WIMPs to annihilate in pairs: This effect is negligible for the point of view of the depletion of dark matter particles (except maybe in the center of very singular dark matter profiles or dark matter clumps) but maybe significant from the point of view of injecting in the hosting galaxy exotic primary cosmic rays ^{7, 8)}.

WIMP annihilations generate quarks, leptons, gauge bosons and Higgs bosons, which in turn fragment and/or decay into stable species we might be able to identify in cosmic ray measurements. Details in annihilation rates and in branching ratios, and hence in the type and abundance of the generated species, depend on the specific particle physics model one considers; e.g., in the minimal supersymmetric extension to the Standard Model (MSSM), the relevant tree-level two-body final states from neutralino annihilations in the halo are: $c\bar{c}$, $b\bar{b}$, $t\bar{t}$, $\tau^+\tau^-$, W^+W^- , Z^0Z^0 , W^+H^- , ZH_1^0 , ZH_2^0 , $H_1^0H_3^0$ and $H_2^0H_3^0$, with branching ratios rather sensitive to the free parameters in the model, but generally with heavy quarks or gauge bosons as dominant channels. Nevertheless, the main features in the process of cosmic ray generations are common to any WIMP model. First, for thermal relic WIMPs, the strength of the total annihilation cross section at zero relative velocity $(v\sigma)_{v=0}$ (the appropriate limit for particles in galactic halos) is not free to vary over several orders of magnitude; in most cases (i.e. far from thresholds and resonances ⁹⁾, and neglecting the case of so called coannihilations ^{9, 10)}) it is expected to be of the same order as for the thermally averaged annihilation cross section, fixed by the WIMP relic abundance, see eq.1. At the same time, it is very hard to define a particle physics model in which the relative strength of the various cosmic ray yields is drastically different from what obtained in another sample case: most particles are not generated promptly, but arise in fragmentation

processes or decay chains, common to all species. This should be kept in mind when testing a given model against a set of cosmic ray data: the comparison with data on other cosmic ray species is always needed and, if the model fits one set of data but severely violates a bound another channel, it is unlikely to find an alternative particle physics model which escapes such bound.

The source function for the species s can be written as:

$$Q_s^x(E, \vec{x}) = (v\sigma)_{v=0} \left(\frac{\rho_\chi(\vec{x})}{m_\chi} \right)^2 \sum_f \frac{dN_s^f}{dE} B^f \quad (2)$$

where for a given annihilation channel f , B^f and dN_s^f/dE are, respectively, the branching ratio and the energy distribution for the yield s . The source function scales with the square of the WIMP number density locally in space, which is equal to the dark matter density divided by the WIMP mass: Any local enhancement in the dark matter density $\rho_\chi(\vec{x})$ (which might be due to the existence of halo profiles singular towards the galactic center or clumps in the halo, as indicated by N-body simulations) corresponds to a sharp increase in the cosmic ray yield, see, e.g. ¹¹). On the other hand, the scaling with the WIMP mass makes the detection of heavy WIMPs harder (a slight suppression for heavy masses is implicitly contained in the annihilation cross section as well).

The stable cosmic ray species generated in WIMP annihilations include gamma-rays, positrons, neutrinos, antiprotons and antideutrons, and, in the same amounts, their counterparts with opposite lepton or baryon numbers. The focus is generally on gamma-rays and antimatter, as in these cases the backgrounds from standard sources are well understood and/or scarce. Neutrino fluxes from WIMP annihilations in the galactic halo are negligible; an exception may be the Galactic center in case a huge density enhancement is present there ¹²) (but see ¹³); recall also that the search for neutrino fluxes from the annihilation of WIMPs that are gravitationally trapped in the center of the Sun or the Earth is another very promising technique to detect WIMP dark matter ¹⁴).

Gamma-rays propagate along geodesics, generally with very little absorption, and hence the flux observed at Earth is just the sum of contributions along a given line of sight. Among proposed dark matter gamma-ray sources, there are the Galactic center, the whole Milky Way halo (in case the dark matter distribution is clumpy), external galaxies ¹⁵) and cosmological sources ¹⁶).

The case for charged particles is slightly more involved as a model for their propagation in the Galactic magnetic fields is needed; with more and more cosmic ray data available, such model is being refined, and it is not any more a source of large uncertainties. By measuring charged cosmic rays one may hope to probe the dark matter distribution in the Milky Way (or, actually, in a nearby portion of it); a more indirect (and model dependent) tool to test the emission of electrons and positrons in WIMP annihilations in far away sources is to search for synchrotron radiation ¹⁷⁾.

The presence of large fluxes providing sizeable excesses is a necessary but not sufficient condition for dark matter detection. In most cases the estimate of the background is not entirely solid; to make a definite statement about the nature of a given excess, excluding more mundane sources, a clear signature for the dark matter induced component is needed. We review below the cases in which such signatures are present; in all other cases the comparison between fluxes and data will just provide upper limits for the exotic component, which in turn can be translated into joint constraints on WIMP models and dark matter distributions.

3 Dark matter signatures in cosmic ray fluxes

The production of antiprotons in WIMP annihilations was the first cosmic ray yield discussed in detail in the literature. There is not standard source of primary antiprotons; those generated in the interaction of primary protons with the interstellar medium, for kinematic reasons, are born with non zero momentum. This suppression at low kinetic energies has no reason to exist for antiprotons from WIMP annihilations: this signature ^{7, 8)} was used to claim detection when early measurements of the low energy antiprotons reported abnormally large fluxes ¹⁸⁾. Such data have not been confirmed by more recent and accurate measurements ^{19, 20)}, now consistent with fluxes of secondary origin alone. Although an exotic component is not ruled, the low energy signature is unfortunately fading out. An alternative signature ²¹⁾, a break in the antiproton spectrum at energies above few tens of GeV, may appear for heavy WIMP dark matter candidates in case the halo of the Milky Way is substantially clumpy. Such break due to a tail of antiprotons produced with very large momenta cannot be mimicked in conventional schemes.

The flux of antideutrons ²²⁾, formed from the merging of an antiproton

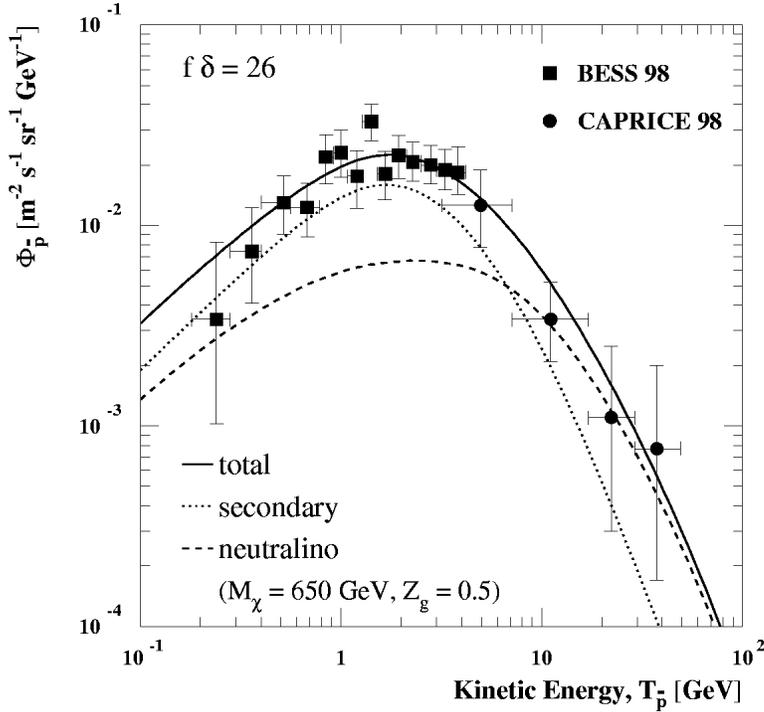


Figure 1: Sample spectrum for the WIMP induced antiproton flux. The dark matter candidate considered here is a neutralino in the anomaly mediated supersymmetry breaking scheme ⁴⁾, with mass equal to 650 GeV and gaugino fraction $Z_g = 0.5$. The dark matter halo is assumed to be clumpy with a moderate clumpyness factor, $f\delta = 26$, see ¹¹⁾ for details.

and an antineutron, is expected to be about four orders of magnitude smaller than for the antiproton flux; nevertheless, the ratio signal to background at low energies may be much higher than for antiprotons. If future instruments achieve the required sensitivities, antideuteron measurements may set severe constraints or, eventually, allow for dark matter detection ^{22, 23)}. A few model dependent assumptions are involved in these conclusions, but this detection method deserves certainly further investigations.

In case gauge bosons or leptons are the dominant channels in WIMP annihilations, the induced positron flux is generated in prompt decays, such as the decay of W^+ and τ^+ ^{24, 25)}, and with a very distinctive spectral signature:

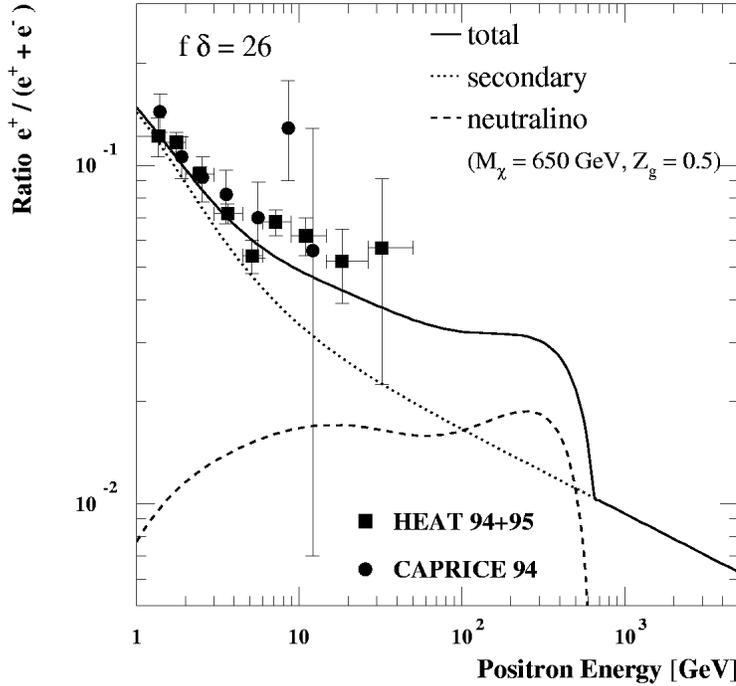


Figure 2: Sample spectrum for the WIMP induced positron flux. The dark matter candidate considered here is the same as in fig.1.

The flux is peaked at an energy equal to about one half of the WIMP mass m_χ , while at higher energies it is sharply suppressed (and it is zero above m_χ). Upcoming experiments ^{26, 27)} will search for such a decrease in the range few tens to few hundreds of GeV. Note, however, that if the positron flux is instead generated in fragmentation processes initiated by WIMP annihilations into quarks, the energy spectrum is shifted to lower energies and becomes rather featureless, definitely more problematic from the point of view of dark matter detection.

Finally, the bulk of the dark matter induced gamma-ray yield arises in the decay of neutral pions produced in fragmentation processes ²⁸⁾. Unfortunately, the π^0 intermediate state is common to other astrophysical processes, and this may turn out to be a severe limiting factor to disentangle dark matter sources. A gamma-ray flux with a much better signature may arise from WIMP

annihilation, at one-loop level, into a two body final state containing a photon: As WIMPs in the galactic halos move with non-relativistic velocities, $v/c \sim 10^{-3}$, the outgoing photons are nearly monochromatic, with energy of the order of the neutralino mass (29, 30, 31). There is no other known astrophysical source with a similar feature: the detection of a line signal out of a spectrally smooth gamma-ray background would be a spectacular confirmation of the existence of dark matter in form of exotic massive particles.

4 Induced fluxes versus available and upcoming measurements

Current cosmic ray data do not show any evidence for an exotic component due to WIMP annihilations. In the energy ranges of interest to search for such an effect, probably the most accurate available measurements are those for antiprotons. At the moment, the antiproton channel is also the one in which the ratio signal to measured background is the highest. Existing data can be used to rule out models for WIMP dark matter, otherwise allowed both by particle physics data and by other astrophysical measurements (32). Data from upcoming experiments (26, 27) should provide even tighter constraints.

An indication of an excess with respect to the standard prediction of the secondary component has been found in positrons measurements (33, 34). This excess opens up a window for a WIMP induced component (35); note however that the ratio signal to background is generally much larger for antiprotons, therefore, typically, if a WIMP signal is indeed present in the positron flux, an exotic component should be singled out in the antiproton flux as well, compare fig.1 with fig.2. A more definite statement will soon be possible, with new data available.

Excesses compatible with radiation from WIMP annihilations have been claimed in gamma-ray measurements as well. Data from the EGRET telescope (energies between few tens of MeV to 20 GeV) seem to show these features in the diffuse gamma-ray flux towards the Galactic center (36) and at high latitudes (37): although an exotic source might account for them, the data may just show that the current model for the diffuse gamma-ray emission from the Galaxy has to be refined. In this case, as in the previous ones, the picture will be soon much clearer, with a much wider set of data that is going to be available. In particular, the next generation of gamma-ray experiments, both ground- and space-based, e.g. (38, 39), will map for the first time the gamma-

ray sky up to energies of hundreds of GeV, in the range where a gamma-line from WIMP annihilations might be found.

In conclusion, we are approaching an exciting time for cosmic ray measurements; although dark matter detection is not the main focus of most of the upcoming experiments, it may turn out to be their greatest achievement.

References

1. M.S. Turner, astro-ph/0106035.
2. G. Jungman, M. Kamionkowski and K. Griest, Phys. Rep. **267**, 195 (1996).
3. W. Hu *et al*, Astrophys. J. **549**, 669 (2001).
4. T. Gherghetta, G.F. Giudice and J.D. Wells, Nucl. Phys. B **559**, 27 (1999).
5. J.F. Navarro, C.S. Frenk and S.D.M. White, Astrophys. J. **462**, 563 (1996).
6. B. Moore *et al*, Astrophys. J. **499**, L5 (1998).
7. J. Silk and M. Srednicki, Phys. Rev. Lett. **50**, 624 (1984).
8. F.W. Stecker, S. Rudaz and T.F. Walsh, Phys. Rev. Lett. **55**, 2622 (1985).
9. K. Griest and D. Seckel, Phys. Rev. D **43**, 3191 (1991).
10. J. Edsjö and P. Gondolo, Phys. Rev. D **56**, 1879 (1997).
11. L. Bergström, J. Edsjö, P. Gondolo and P. Ullio Phys. Rev. D **59**, 043506 (1999).
12. P. Gondolo and J. Silk, Phys. Rev. Lett. **83**, 1719 (1999).
13. P. Ullio, H.S. Zhao and M. Kamionkowski, Phys. Rev. D **64**, 043504 (2001).
14. J. Silk, K. Olive and M. Srednicki, Phys. Rev. Lett. **55**, 257 (1985).
15. E.A. Baltz *et al*, Phys. Rev. D **61**, 023514 (2000).
16. L. Bergström, J. Edsjö and P. Ullio, Phys. Rev. Lett. in press, astro-ph/0105048.
17. S. Colafrancesco and B. Mele, astro-ph/0008127.

18. A. Buffington *et al*, *Astrophys. J.* **248**, 1179 (1981).
19. T. Sanuki *et al*, *Astrophys. J.* **545**, 1135 (2000).
20. M. Boezio *et al*, astro-ph/0103513.
21. P. Ullio, PhD thesis, Physics Department, Stockholm University, (1999).
22. F. Donato, N. Fornengo and P. Salati, *Phys. Rev. D* **62**, 043003 (2000).
23. K. Mori *et al*, astro-ph/0109463.
24. M. Kamionkowski and M.S. Turner, *Phys. Rev. D* **43**, 1774 (1991).
25. E.A. Baltz and J. Edsjö, *Phys. Rev. D* **59**, 023511 (1999).
26. O. Adriani *et al*, Proc. of the 26th ICRC, Salt Lake City, OG.4.2.04 (1999).
27. S. Ahlen *et al*, *Nucl. Instrum. Methods A* **350** 351 (1994).
28. J. Silk and H. Bloemen, *Astrophys. J.* **313**, L47 (1987).
29. L. Bergström and H. Snellman, *Phys. Rev. D* **37**, 3737 (1988).
30. L. Bergström and P. Ullio, *Nucl. Phys. B* **504**, 27 (1997).
31. P. Ullio and L. Bergström, *Phys. Rev. D* **57**, 1962 (1998).
32. P. Ullio, *JHEP* **0106**, 053 (2001).
33. S.W. Barwick *et al*, *Astrophys. J.* **482**, L191 (1997).
34. M. Boezio *et al*, *Astrophys. J.* **532**, 653 (2000).
35. G.L. Kane, L.T. Wang and J.D. Wells, hep-ph/0108138.
36. H.A. Mayer-Hasselwander *et al*, *Astron. Astrophys.* **335**, 161 (1998).
37. D.D. Dixon *et al*, *New Astron.* **3**, 539 (1998).
38. M. Martinez *et al*, Proc. of the 26th ICRC, Salt Lake City, OG.4.3.08 (1999).
39. Gamma-ray Large Area Space Telescope (GLAST), homepage <http://www-glast.stanford.edu>.

**MSSM parameters space regions allowed by indirect neutralino
detection**

Andrea Lionetto
*Dipartimento di Fisica Università Roma Tor Vergata
Viale della Ricerca Scientifica, 1 00173 Roma, Italy*

ABSTRACT

We show a simple method, based on the standard χ^2 test statistic, to discriminate what supersymmetric models are detectable by indirect searches of neutralino annihilation. We can then put different constraints on the allowed region of the parameters space of the MSSM.

1 Introduction

The neutralino seems to be one of the most promising candidate as a constituent for cold dark matter ¹⁾, i.e. non relativistic at the temperature of the freeze out. It appears to be the LSP (Lightest Supersymmetric Particle) in a large portion of the parameters space of the MSSM, that is the minimal supersymmetric extension of the standard model with R-parity conservation

4). Indirect detection is possible through the search of neutralino annihilation products that are produced in the galactic halo. The detectable products are the typical constituents of cosmic rays such as antiprotons, positrons and gamma rays ²⁾. Here we consider mainly the antiprotons production ³⁾. Cosmic ray induced antiprotons are generated mainly through $pp \rightarrow \bar{p} + X$ collisions of cosmic ray protons with interstellar medium.

2 Theoretical framework

We work in the framework of the Minimal Supersymmetric extension of the Standard Model (MSSM) with the allowed renormalizable soft supersymmetry breaking terms ⁴⁾. The phenomenological parameters are reduced to 7 starting from the usual 125 ⁵⁾: μ the higgsino mass parameter, M_2 the gaugino mass parameter, m_A the CP-odd Higgs bosons mass, $\tan(\beta) = v_2/v_1$ the higgs bosons vacuum expectation ratio, m_q the scalar mass parameter, A_b and A_t the trilinear coupling in the bottom and top sector respectively.

These parameters are given at the electroweak symmetry breaking scale, i.e. at energy of order of 1 TeV.

The lightest neutralino is defined, as usual, as a linear combination of the gauginos \tilde{B} and \tilde{W}^3 , the superpartners of the $U(1)$ gauge field B and the third component of the $SU(2)$ gauge field W^3 that appears in the standard model, and the neutral higgsinos \tilde{H}_1^0 and \tilde{H}_2^0 , that are the superpartners of two of the Higgs bosons doublet neutral components that appears explicitly in the MSSM:

$$\chi = N_{10}\tilde{B} + N_{20}\tilde{W}^3 + N_{30}\tilde{H}_1^0 + N_{40}\tilde{H}_2^0 \quad (1)$$

We have used the DarkSUSY ⁶⁾ fortran routines for the differential antiproton flux calculation. We have considered, at tree level, the relevant states for \bar{p} production. These include all the heavier quarks (c, b, and t), gauge bosons and Higgs bosons, and the subsequent hadronization of the states. In input we have used a random generated sampling of the parameters space of the order of 10^5 models. The only constraint used here, in addition to the physical consistency, is that on the neutralino relic density, in order to get rid of this dependence. We have used for the allowed range for the cold dark matter relic abundance:

$$0.1 \leq \Omega_{DM} h^2 \leq 0.3 \quad (2)$$

where Ω_{DM} is the ratio between the dark matter density and the critical density and h is a parameter in the Hubble constant.

3 Supersymmetric detectable models

The problem is to discriminate between different supersymmetric models. The intuitive idea is to “calculate” how far are the two flux curves, the background and the supersymmetric contribution (fig. 1). The first curve, the background curve, is obtained considering only the standard model production and propagation of the \bar{p} [7]. In the background contribution is not present any source of exotic components. The other curve, the supersymmetric contribution, represents the contribution coming from a source of annihilating neutralinos.

To do this job we can use an hypothesis test method. The simplest of this method is the χ^2 test. We use the usual definition of the reduced χ^2 as:

$$\chi_r^2 = \frac{1}{N-1} \sum \frac{(y_i - x_i)^2}{\sigma_i^2} \quad (3)$$

where N is the number of degrees of freedom, i.e. the number of points in which we calculate the fluxes. In order to be able to apply this kind of test we must associate an error to the points of, at least, one of the theoretical flux curves (the standard one or the standard plus susy contribution). A simplifying assumption is to consider only the statistical errors naturally coming from a counting experiment. So, indicating with N_i the number of counts, the associated error is the Poisson one:

$$\sigma_i = \sqrt{N_i} \quad (4)$$

There is a simple relation between flux and number of counts, that depends explicitly by the detector characteristics of the particular experiment we want to consider:

$$N_i = \phi_{\bar{p}} \cdot A \cdot \Delta t \cdot \Delta E_i \quad (5)$$

where $\phi_{\bar{p}}$ is the antiproton flux in units of $GeV^{-1}m^2sr^{-1}s^{-1}$, A is the effective area of the detector, Δt is the acquisition data time and ΔE_i is the

energy bin. With the aim of this formula it's straightforward to calculate the statistical error associated to the flux (fig. 1), that is:

$$\Delta\phi = A \cdot \Delta t \cdot \Delta E_i \cdot \sigma_i \quad (6)$$

Fixing the significance level, we accept, as detectable, all the models, for a given number of degrees of freedom, that satisfy:

$$\chi_r^2 \geq c \quad (7)$$

where c can be obtained from the standard tabulation ⁸⁾, i.e. integrating the probability distribution function of the χ_r^2 .

4 Limits on the parameters space

We can use this machinery in order to single out parameters space regions that generate models with detectable flux. This can be achieved for example, making contour plots of χ_r^2 vs. model parameters. The main problem is to reduce the number of parameters to consider. There are two ways to do this. One is to start *ab initio* with a more constrained models, such as for example the CMSSM, or models with anomaly mediated supersymmetry breaking. This is, by far, the most appealing possibility from a pure theoretical ground, because at the end we want some theory with less phenomenological parameters as possible. The other is to identify the relevant parameters for a given process, fixing the values of the others. This can be justified if the χ_r^2 weakly depends by some parameters.

When we consider the processes of \bar{p} production (fig. 2), we see that two of the relevant parameters are the neutralino mass M_χ and the gaugino fraction Z_g , that are function, in terms of the fundamental parameters that define the model, of M_2 , μ and $\tan(\beta)$.

5 Conclusions

The method, described above, allows to understand what kind of models are phenomenologically interesting, considering only indirect detection of neutralino annihilations. This implies that we are able to identify what regions of parameters space can really be probed, and what regions are not yet explorable. We

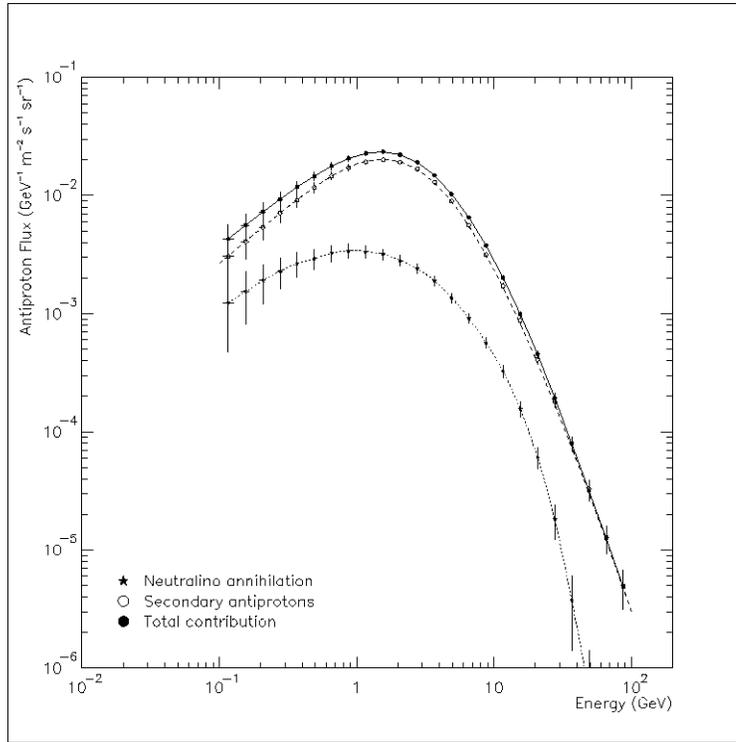


Figure 1: *flux from a particular supersymmetric model.*

can also consider immediate extension for positrons and gamma rays, in order to put more constraints.

One possible development concerns the possibility to apply some other test hypothesis method, such as the generalized likelihood ratio, or even some more sophisticated method such as bayesian and neural networks. It would be interesting to find some algorithmic procedure in order to be able to scan a large portion of the parameters space.

The other possible development concerns the study of theoretical models that involve a less number of parameters, such as the anomaly mediated

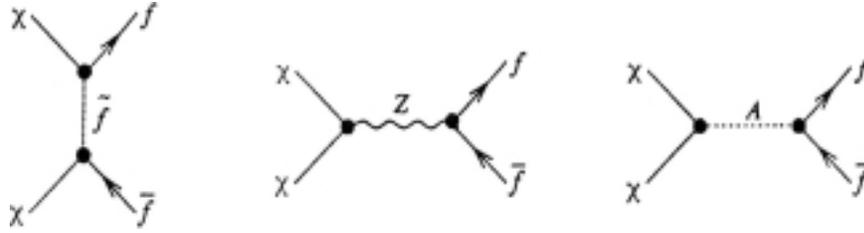


Figure 2: *relevant diagrams or the process $\chi\chi \rightarrow p\bar{p}$.*

supersymmetry breaking models.

References

1. G. Jungman *et al*, “Supersymmetric Dark Matter”, Phys. Rept. **267**, 195 (1996)
2. L. Bergstrom *et al*, Phys. Rev. D **59**, 043506 (1999)
3. L. Bergstrom *et al*, Astrophysical Journal, **526**, 215 (1999)
4. S.P. Martin, “A Supersymmetry Primer”, hep-ph/9709356
5. L. Bergstrom *et al*, Astroparticle Physics, **5**, 263 (1996)
6. P. Gondolo *et al*, astro-ph/0012234
7. A.W. Strong *et al*, Astrophysical Journal, **509**, 212 (1998)
8. G. Cowan, “Statistical Data Analysis”, (Oxford University Press, Oxford, 1998)

The engines of the Universe

M.Vietri	What have we learned about gamma ray bursts from afterglows?
A.Dar	The threat to life from Eta Carinae and gamma-ray bursts
F.Longo	Gamma-ray satellites and gamma-ray bursts
N.Omodei	Emission Model and GRB Simulations
C. Aramo	AUGER observatory: the world's largest cosmic ray detector
A.Bettini	(New) Neutrino Physics at Gran Sasso Laboratory
J.Primack	Probing galaxy formation with high energy gamma-rays

WHAT HAVE WE LEARNED ABOUT GAMMA RAY BURSTS FROM AFTERGLOWS?

Mario Vietri
Università di Roma 3

ABSTRACT

The discovery of GRBs' afterglows has allowed us to establish several facts: their distance and energy scales, the fact that they are due to explosions, that the explosions are relativistic, and that the afterglow emission mechanism is synchrotron radiation. On the other hand, recent data have shown that the fireball model is wrong when it comes to the emission mechanism of the true burst (which is unlikely to be synchrotron again) and that shocks are not external. Besides these relatively tame points, I will also discuss the less well established physics of the energy deposition mechanism, as well as the possible burst progenitors.

1 Introduction

Gamma ray bursts (GRBs) were discovered in 1969 (Klebesadel, Strong and Olson 1973) by American satellites of the *Vela* class aimed at verifying Russian

compliance with the nuclear atmospheric test ban treaty. Though the discovery was made in 1969, the paper appeared only four years later because the authors had lingering doubts about the reality of the effects they had discovered. Since then, several thousands of bursts have been observed by a more than a dozen different satellites, but it is remarkable that the basic burst features outlined in the abstract of the 1969 paper (photons in the range $0.2 - 1.5 \text{ MeV}$, durations of $0.1 - 30 \text{ s}$, fluences in the range $10^{-5} - 2 \times 10^{-4} \text{ ergs cm}^{-2}$) have remained substantially unchanged.

Current evidence (Fishman and Meegan 1995) has highlighted a wide ($0.01 - 100 \text{ s}$) duration distribution, with hints of a bimodality which is claimed to correlate (at the 2.5σ level) with spectral properties. All bursts' spectra observed so far are strictly non-thermal, and there has never been any confirmation by BATSE of a supposed thermal component (nor of cyclotron lines or precursors, for this matter) claimed in previous reports. A remarkable feature reported by BATSE is the bewildering diversity of light curves, ranging from impulsive ones (a spike followed by a slower decay, nicknamed FREDs for Fast Rise-Exponential Decay), to smooth ones, to long ones with amazingly sharp fluctuations, including even some with a strongly periodic appearance (two such examples are the 'hand' and the 'comb', so nicknamed from the number of high-Q, regularly repeating sharp spikes).

The most exceptional result from BATSE, though, was the sky distribution of the bursts (Fig.1). It was obvious from it that the bursts *had* to be extragalactic, as already discussed by theorists (Usov and Chibisov 1975, Paczyński 1986).

2 Afterglows

The next major step was triggered by BeppoSAX: in the summer of 1996, L. Piro and his coworkers located in archival data of the satellite the soft X-ray counterpart of a GRB (GRB 960720). They immediately conceived the idea of implementing a procedure to follow the next burst in real time, by re-orienting the whole satellite, after the initial detection by the Wide Field Cameras, so that the more sensitive Narrow Field Instruments could pinpoint the burst location to within 45 arcsecs, a feat never achieved in such short times, and by a single satellite. After an initial snafu (GRB 970111), the gigantic effort paid off with the discovery of the X-ray afterglow of GRB 970228 (Costa *et al.*,

2000 BATSE Gamma-Ray Bursts

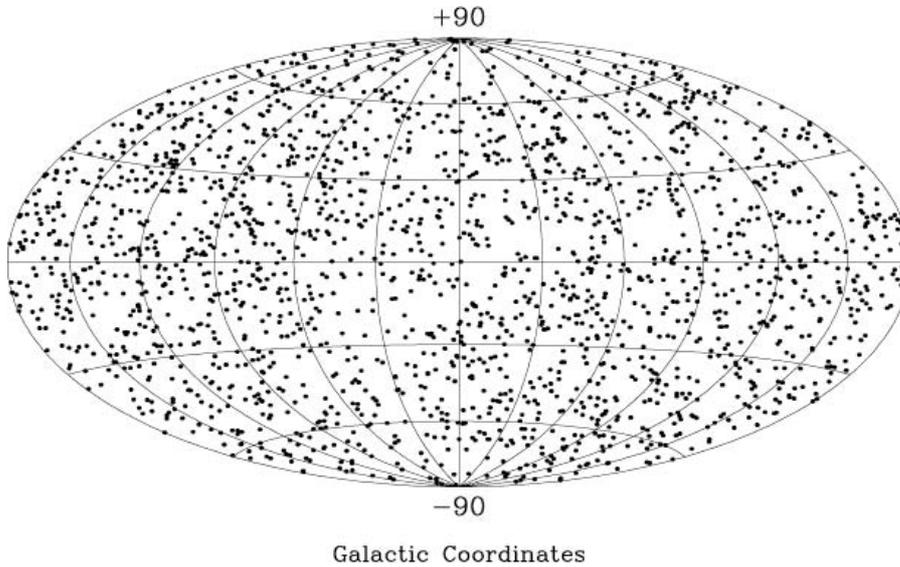


Figure 1: *Burst distribution on the plane of the sky*

1997), immediately followed by the discovery of its fading optical counterpart (van Paradijs *et al.*, 1997), obtained through a search inside the WFC error box, in perfect agreement with theoretical predictions (Vietri 1997a, Mészáros and Rees 1997).

After the detection of the optical counterpart, the door was open to find the bursts' redshifts: Table I summarizes the status of our current knowledge (September 1999); bursts' luminosities are for isotropic sources. Two comments are in order. First, the bursts have *prima facie* a redshift distribution not unlike that of AGNs and of the Star Formation Rate (SFR). The initial hope that they might trace an even more distant and elusive Pop III, triggered by the fact that the second redshift detected was also the largest so far (GRB 971214, $z = 3.4$), has now vanished. Second, in order to place the energy release of GRB 990123 in context, one should notice that 4×10^{54} ergs is the energy obtained by converting the rest-mass of two solar masses, or, alternatively, the energy emitted by the whole Universe out to $z \approx 1$ within the burst duration.

GRB	z	E_{iso}
970228	0.695	$5 \times 10^{51} \text{ erg}$
970508	0.835	$2 \times 10^{51} \text{ erg}$
971214	3.4	$3 \times 10^{53} \text{ erg}$
980703	0.93	$3 \times 10^{53} \text{ erg}$
990123	1.7	$4 \times 10^{54} \text{ erg}$
990510	1.6	$2 \times 10^{53} \text{ erg}$
990712	0.43	
991208	0.70	$1.3 \times 10^{53} \text{ erg}$
991216	≥ 1.02	

So, a single (perhaps double) star outshines the whole Universe.

Besides the distance and energy scales, the major impact of the discovery of afterglows has been the establishment of some key features of the fireball model (Rees and Mészáros 1992):

1. bursts are due to explosions, as evidenced by their power-law behaviour;
2. the explosions are relativistic, as proved by the disappearance of radio flares;
3. the burst emission is due to synchrotron emission, as shown by the afterglow spectrum, and its optical polarization.

I will illustrate these points in the following, but, lest we become too proud, we should also remember that the fireball model has met some failures. The original version of the model (Mészáros and Rees 1993) advocated the dissipation of the explosion energy at external shocks (*i.e.*, those with the interstellar medium). Sari and Piran (1997), following a point originally made by Ruderman (1975) showed that these shocks smooth out millisecond timescale variability, which can only be maintained by the internal shocks proposed by Paczyński and Xu (1994). Also, the fireball model originally ascribed even the emission from the burst proper (as opposed to the afterglow) to optically thin synchrotron processes; I will discuss in the section *Embarrassments* why this is exceedingly unlikely. Furthermore, even the last tenet of mid-90s common wisdom, *i.e.*, that bursts are due to neutron binary mergers, does not look too promising at the moment (since some bursts seem to be located inside star

forming regions, incompatible with the long spiral-in time), though of course it is by no means ruled out yet.

2.1 The fireball model

Here, one may assume that an unknown agent deposits $10^{51} - 10^{54}$ *ergs* inside a small volume of linear dimension $\approx 10^6 - 10^7$ *cm*. The resulting typical energy density corresponds to a temperature of a few *MeVs*, so that electrons and positrons cannot be bound by any known gravitational field. In these conditions, optical depths for all known processes exceed 10^{10} . The fluid expands because of its purely thermal pressure, converting internal into bulk kinetic energy. Parametrizing the baryon component mass as $M_b \equiv E/\eta c^2$, it can be shown that, for $1 \leq \eta \leq 3 \times 10^5$ (Mészáros, Laguna, Rees 1993) the fluid achieves quickly (the fluid Lorentz factor increases as $\gamma \propto r$) a coasting Lorentz factor of $\gamma \approx \eta$.

The requisite asymptotic Lorentz factor is dictated by observations: photons up to $\epsilon_{ex} \approx 18$ *GeV* have been observed by EGRET from bursts (Fishman and Meegan 1995). For these photons to evade collisions with other photons, and thus electron/positron pair production, it is necessary that, in the reference frame in which a typical burst photon (with $\epsilon \approx 1$ *MeV*) and the exceptional photon are emitted, they appear as below pair production threshold: thus we must have $\epsilon' \epsilon'_{ex} \leq 2m_e c^2$. Since $\epsilon' \approx \epsilon/\gamma$, and similarly for the other photons, we find (Baring 1993)

$$\gamma \approx 300 \left(\frac{\epsilon}{1 \text{ MeV}} \frac{\epsilon_{ex}}{10 \text{ GeV}} \right)^{1/2}. \quad (1)$$

From what we said above, we thus require a maximum baryon contamination, in an explosion of energy E , of $M_b < 10^{-6} M_\odot (E/10^{51} \text{ erg})(300/\eta)$.

The energy release is now assumed to be in the form of an inhomogeneous wind, with parts having a Lorentz factor larger than parts emitted previously. This leads to shell collisions (the internal shock model) at radii r_{sh} which allow a time-scale variability $\delta t \approx r_{sh}/2\gamma^2 c$; for $\delta t = 1$ *ms*, $r_{sh} \approx 10^{13}$ *cm*, which fixes the internal shock radii. Particle acceleration at these internal shocks and ensuing non-thermal emission is thought to lead to the formation of the burst proper. At larger radii, a shock with the surrounding ISM forms, and shell deceleration begins at a radius $r_{ag} = (3E/4\pi n m_p c^2 \gamma^2)^{1/3} \approx 10^{17}$ *cm* for a $n = 1 \text{ cm}^{-3}$ particle density typical of galactic disks. It is thought that the

afterglow begins when the shell begins the slowdown, as this drives a marginally relativistic shock into the ejecta, thusly extracting a further fraction of their bulk kinetic energy.

2.2 Why explosions

The success of the fireball model lies in this, that it decouples the problem of the energy injection mechanism from the following evolution, which is, furthermore, an essentially hydrodynamical problem. It can be shown, in fact (Waxman 1997) that the evolution of the external shock is adiabatic, that the shock Lorentz factor decreases as $\gamma \propto r^{-3/2}$ because of the inertia of the swept-up matter, and thus r scales with observer's time as $t = r/\gamma^2 c \rightarrow \gamma \propto t^{-3/8}$ (for a radiative solution $\gamma \propto r^{-3/7}$, Vietri 1997b). If afterglow emission is due to optically thin synchrotron in a magnetic field in near-equipartition with post-shock energy density, it can be shown that $B \propto \gamma$, that the typical synchrotron frequency at the spectral peak $\nu_m \propto \gamma B \gamma_e^2 \propto \gamma^4$ (where $\gamma_e \propto \gamma$ is the lowest post-shock electron Lorentz factor), and that $F(\nu_m) \propto t^{-3\beta/2}$, where β is the afterglow spectral slope. As it can be seen, these expectations are based exclusively upon the hydrodynamical evolution (and the synchrotron spectrum), and are thus reasonably robust.

We thus expect power-law time decays, a characteristic of strong explosions (see the Sedov analogy!), with time- and spectral-indices closely related. This is what is observed everywhere, from the X-ray through the optical to the radio, (see Piro and Fruchter, this volume), the few exceptions being discussed later on. In fact, the equality of the time-decay index of the X-ray and optical data in afterglows of individual sources has been taken as the key element to show that emission in the different bands is due to the same source. Time indices in the X-ray are in the range 0.7 – 2.2 (Frontera *et al.*, 2000).

2.3 Why synchrotron spectrum in the afterglow

After having established that bursts are due to explosions, we happily learn that afterglows emit through synchrotron processes. In fig. 2 (Galama *et al.*, 1998), we show the superposition of theoretical expectations for an optically thin synchrotron spectrum (including the cooling break at $\nu \approx 10^{14}$ Hz) with observations for GRB 970508. The remarkable agreement is even more exciting as we remark that observations are not truly simultaneous, but are scaled back

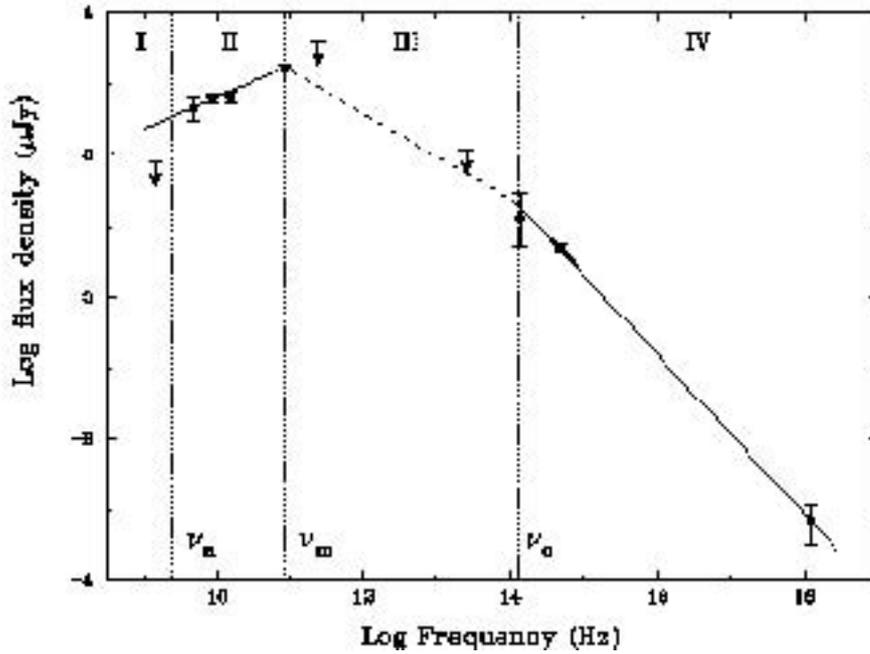


Figure 2: *Simultaneous spectrum of the afterglow of GRB 970508, from Galama et al., 1998.*

to the same time by means of the theoretically expected laws for time-decay, thus simultaneously testing the correctness of our hydro. Another piece of evidence comes from the discovery of polarization in the optical afterglow of GRB 990510 (Fig. 3, Covino *et al.*, 1999, Wijers *et al.*, 1999). This polarization may appear small ($\approx 2\%$), but it is surely not due to Galactic effects: stars in the same field show a comparable degree of polarization, but along an axis different by about 50° . Also, polarization in the source galaxy is unlikely, because of a very stringent upper limit on the reddening due to this galaxy (Covino *et al.*, 1999). The only remaining question mark is emission from an anisotropic source, but this would require a disk of 10^{18} cm to survive the intense γ ray (and X, and UV) flash: though not excluded, it does not look likely.

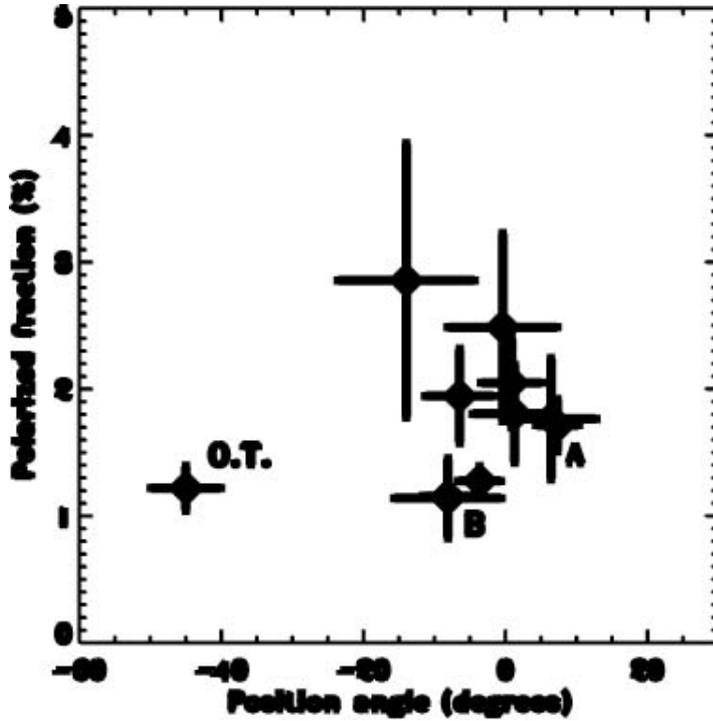


Figure 3: Polarization amplitude and position angle for optical afterglow of GRB 990510, from Covino *et al.*, 1999.

2.4 Why relativistic expansion

Radio observations of the first burst observed so far (GRB 970508, Frail *et al.*, 1997) showed puzzling fluctuations by about a factor of 2 in the flux, over a time-scale of days, disappearing after about 30 days from the burst (Fig. 4). This extreme, and unique behaviour, was explained by Goodman (1997), who showed that it is due to interference of rays travelling along different paths through the ISM, and randomly deflected by the spatially varying refractive index of the turbulent ISM. The wonderful upshot of this otherwise marginal phenomenon, is that these effects cease whenever the source expands beyond a radius

$$R = 10^{17} \text{ cm} \frac{\nu_{10}^{6/5}}{d_{sc,kpc} h_{75}} \left(\frac{SM}{10^{-2.5} m^{-20/3} \text{ kpc}} \right)^{-3/5}, \quad (2)$$

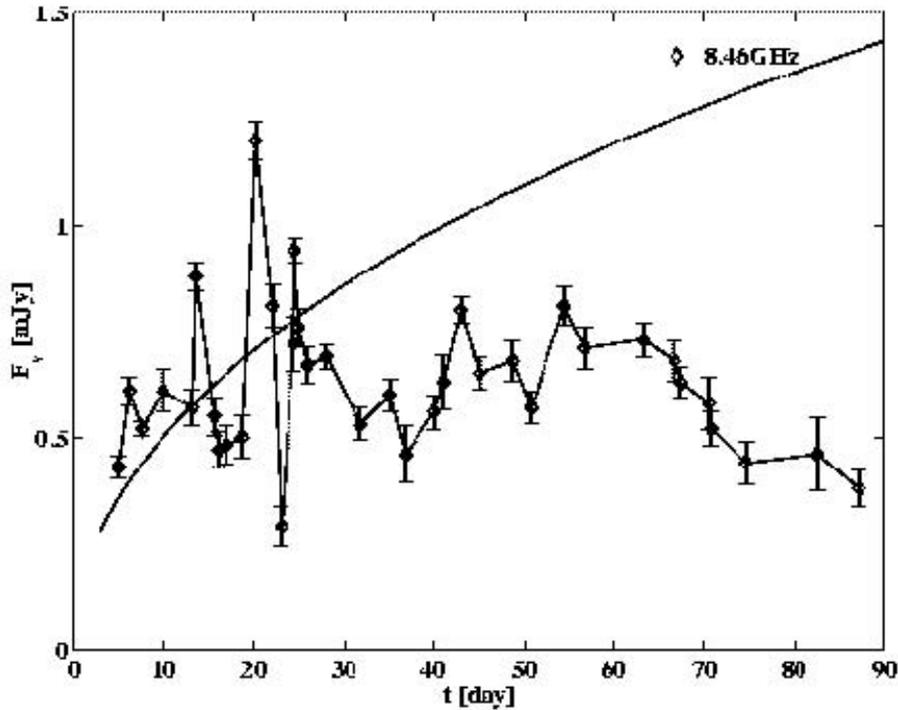


Figure 4: VLA observations at 8.46 GHz of the afterglow of GRB 970508, from Waxman, Frail and Kulkarni 1998.

where ν_{10} is the radio observing frequency in units of 10^{10} Hz, $d_{sc,kpc}$ is the distance of the ISM from us (assumed to be a uniform scattering screen), and SM is the Galactic scattering measure, scaled to a typical Galactic value. The existence of interference effects is made more convincing by the amplitude of the average increase (a factor of 2, as observed), the correctness in the prediction of the time-interval between different peaks, and of the decorrelation bandwidth. Since flares disappear after about 30 days, it means that the average speed of the radio source is $R/30days = 3 \times 10^{10} \text{ cm s}^{-1}$. So we see directly that GRB 970508 expanded at an average speed of c over a whole month, giving us a direct observational proof that the source is highly relativistic. This proof is

completely equivalent to superluminal motions in blazars, and is the strongest evidence in favor of the fireball model.

2.5 GRB 970508: our best case

The afterglow of GRB 970508 is our best case so far: it is in fact a burst for which not only do we know the redshift, but also a radio source that has been monitored for more than 400 days after the explosion (Frail, Waxman and Kulkarni 2000). Through these observations we can see the transition to a sub-relativistic regime at $t \approx 100 d$, measure the total energetics of the following Sedov phase (unencumbered by relativistic effects!) $E_{New} = 5 \times 10^{50} \text{ ergs}$, determine two elusive parameters, $\epsilon_{eq} = 0.5$ and $\epsilon_B = 0.5$ (the efficiencies with which energy is transferred to post-shock electrons by protons, and with which an equipartition field is built up), and the density of the surrounding medium $n \approx 0.4 \text{ cm}^{-3}$. All of these values look reasonable (perhaps ϵ_{eq} and ϵ_B exceed our expectations by a factor of 10, a fact that could be remedied by introducing a slight density gradient which would keep the shock more efficient), so that our confidence in the external-shock-in-the-ISM model is boosted.

Another precious consequence of these late-time observations is that they yield information on beaming and energetics. In fact, GRB 970508 appeared to have a kinetic energy of $E_{rel} = 5 \times 10^{51} \text{ erg}$ when in the relativistic phase, a measurement which can be reconciled with E_{New} (remember that the expansion is adiabatic, so that we must have $E_{New} = E_{rel}$!) only if the unknown beaming angle, assumed $= 4\pi$ in deriving E_{rel} , is smaller than 4π by the factor E_{New}/E_{rel} ; we thus have the only measurement of $\delta\Omega/4\pi = 0.1$, so far. This already rules out all classes of models requiring unplausibly large amounts of beaming, 10^{-8} or even beyond. Hopefully, more such measurements will come in the future, since this observationally heavy method is subject to many fewer uncertainties than the competing method of trying to locate breaks in the time-decay of afterglows. Also, the radiative efficiency of the burst can be estimated: correcting the observed burst energy release $E_{GRB} = 2 \times 10^{51} \text{ erg}$ for the same beaming factor, the radiative efficiency is $E_{GRB}\delta\Omega/4\pi/(E_{New} + E_{rel}\delta\Omega/4\pi) = 0.3$, again a unique determination. Notice however that this figure is subject to a systematic uncertainty: we do not know whether the beaming fraction is the same for the burst proper and for the afterglow.

3 Embarrassments

Something is rotten in the fireball kingdom as well, namely, departures from pure power-law behaviours, and the spectra of the bursts proper.

3.1 Unpowerlawness

Departures from power-laws are expected when one considers the extremely idealized character of the solutions discussed so far: perfect spherical symmetry, uniform surrounding medium, smooth wind from the explosion, ϵ_{eq} and ϵ_B constant in space and time. The tricky point here is to disentangle these distinct factors. In GRB 970508 and GRB 970828 (Piro *et al.*, 1999, Yoshida *et al.*, 1999) a major departure was observed in the X-ray emission, within a couple of days from the burst; they constitute the single, largest violations observed so far, in terms of number of photons. It is remarkable that spectral variations were simultaneously observed, and that both bursts showed traces (at the 2.7σ significance level) of an iron emission line. The similarity of the bursts' behaviour argues in favor of the reality of these spectral features, which have been interpreted as thermal emission from a surrounding stellar-size left-over, pre-expelled by the burst's progenitor (Lazzati *et al.*, 1999, Vietri *et al.*, 1999). Clearly, these departures hold major pieces of information on the bursts' surroundings, and the nature of bursts' progenitors.

It has been argued (Rhoads 1997) that, whenever the afterglow shell decelerates to below $\gamma \approx 1/\theta$, where θ is the beam semi-opening angle, emission should decrease because of the lack of emitting surface, compared to an isotropic source. But, in view of the existence of clear environmental effects (GRB 970508 and GRB 970828), it appears premature to put much stock in the interpretation of time-power-law breaks as due to beaming effects. And equally, it appears to this reviewer that the same comment applies to the interpretation of a resurgence of flux as due to the appearance of a SN remnant behind the shell. The major uncertainty here is the non-uniqueness of the interpretation: Waxman and Draine (2000) have shown that effects due to dust can mimic the same phenomenon.

3.2 Bursts' spectra

A clear prediction of the emission of optically thin synchrotron is that the low-photon-energy spectra should scale like $dN_\nu/d\nu \propto \nu^\alpha$, with $\alpha = -3/2$, since the emission is in the fast cooling regime. Within thin synchrotron, there is no way to obtain $\alpha > -3/2$. This early-recognized requirement (Katz 1994) is so inescapable that it has been dubbed the 'line of death'. Observations are notoriously discordant with this prediction. Preece *et al.* (1999) have shown that, for more than 1000 bursts, α is distributed like a bell between -2 and 0 , with mean $\bar{\alpha} \approx -1$. The tail of this distribution also contains a few tens of objects with $\alpha \approx +1$. An example of these can be found in Frontera *et al.*, 1999 (GRB 970111), which is instructive since BeppoSAX has better coverage of the critical, low-photon-energy region. In particular, BATSE seems to lose sensitivity below ≈ 30 keV, but this is still not enough to explain away the discrepancy with the theory. Also, Preece *et al.*, 1999, showed that the time-integrated spectral energy distribution has a peak at a photon energy $\epsilon_{pk} \approx 200$ keV, and that ϵ_{pk} has a very small variance from burst to burst. Again, this does not seem dependent upon BATSE's lack of sensitivity above 700 keV, and again this has no explanation within the classic fireball model.

Many people agree that the neglect of Inverse Compton processes may be the root of the disagreement: the trick here is to devise a fireball model that smoothly incorporates them. One should remember that the details of the fireball evolution are *generic, i.e.*, they do not depend upon any detailed property of the source, so that things like the radius at which the fireball becomes optically thin (to pairs or baryonic electrons), the radius at which acceleration ends, the equipartition magnetic field, and so on, are all reliably and inescapably fixed by the outflow's global properties. A step toward the solution has been made by Ghisellini and Celotti (1999) who remarked that at least some bursts have compactness parameters $l = 10(L/10^{53} \text{ erg s}^{-1})(300/\gamma)^5 \gg 1$. Under these conditions, a pair plasma will form, nearly thermalized at $kT \approx m_e c^2$, and with Thomson optical depth $\tau_T \approx 10$. The modifications which this plasma will bring to the burst's spectrum are currently unknown, but it may be remarked that this configuration will be optically thick to both high-energy synchrotron photons due to non-thermal electrons accelerated at the internal shocks, and to low-energy cyclotron photons emitted by the thermal plasma, but it will be optically thin in the intermediate region reached by cyclotron photons up-

scattered via IC processes off non-thermal electrons. A model along this line (*i.e.*, upscattering of cyclotron photons by highly relativistic electrons) is in preparation, but it remains to be seen whether it (like any other model, of course) can simultaneously explain the spectral shape and the narrow range of the spectral distribution peak energy ϵ_{pk} .

4 On the central engine

As remarked several times already, the fireball evolution is independent of the source nature. The only existing constraint is the maximum amount of baryon contamination, which is

$$M_b = \frac{E}{\eta c^2} = 10^{-6} M_\odot \frac{E}{10^{51} \text{ erg}} \frac{300}{\gamma}. \quad (3)$$

This is a remarkably small value: since the inferred luminosities exceed the Eddington luminosity by 13 orders of magnitude, they clearly have all it takes to disrupt a whole star, no matter how compact. Yet, the energy deposition must somehow occur outside the main mass, lest the explosion be slowed down to less relativistic, or even possibly Newtonian speeds. In order to satisfy this constraint, it has emerged that the most favorable configuration has a stellar-mass black hole ($M_{BH} \approx 3 - 10 M_\odot$) surrounded by a thick torus of matter ($M_t \approx 0.01 - 1 M_\odot$, with $\rho \approx 10^{10} \text{ g cm}^{-3}$). The presence of a black hole is *not* required by observations in any way: models involving neutron stars are still admissible, the advantage of having a black hole being only the deeper potential well: you may get more energy out per unit accreted mass. The configuration thusly envisaged has a cone surrounding the symmetry axis devoid of baryons, since all models leading to this configuration have large amounts of specific angular momentum, and thus baryons close to the rotation axis either are not there, or have accreted onto the black hole due to their lack of centrifugal support.

4.1 Energy release mechanism

There are two major mechanisms for energy release discussed in the literature, the first one to be proposed (Berezinsky and Prilutskii 1986) being the reaction $\nu + \bar{\nu} \rightarrow e^- + e^+$. Neutrinos have non-negligible mean free paths in the tori envisaged here, so that this annihilation reaction will take place not inside tori

themselves, where they are preferentially generated because densities are highest, but in a larger volume surrounding the source. This is both a blessing and a disgrace: by occupying a larger volume, the probability that every neutrino finds its antiparticle to annihilate decreases, but then the energy is released in baryon-cleaner environments. The problem, though complex, is eminently suitable for numerical simulations, showing (Janka *et al.*, 1999, and references therein) that about 10^{50} ergs can be released this way, above the poles of a black hole where less than $10^{-5}M_{\odot}$ are found.

Highly energetic bursts cannot be reproduced by this mechanism, due to its low efficiency: the second mechanism proposed involves the conversion of Poynting flux into a magnetized wind. The basic physical mechanisms are well-known (Usov 1992) since they have been studied in the context of pulsar emission: electrons are accelerated by a motional electric field $\vec{E} = \vec{v} \wedge \vec{B}/c$ due to the rotation of a sufficiently strong magnetic dipole, attached either to a black hole, or to the torus. Photons are then produced by synchrotron or curvature radiation, and photon/photon collisions produce pairs, to close the circle and allow looping. In order to carry away 10^{51} erg s^{-1} , a magnetic field of $\approx 10^{15}$ G is required. This is not excessive, since it is about three orders of magnitude below equipartition with torus matter, and because such fields already exist in nature, see SGR 1806-20 and SGR 1900+140: the key point is to understand whether some kind of dynamo effect can lead to these high values within the short allotted time.

Depending upon whether the open magnetic field lines extending to infinity are connected to the black hole or to the torus, the source of the energy of the outflow will be the rotational energy of the black hole (the so-called Blandford-Znajek effect) or of the torus. The first case is traditionally discussed in the context of AGNs (Rees, Blandford, Begelman and Phinney 1984), but it is harshly disputed whether the energy outflow may be actually dominated by the black hole rather than by the disk (Ghosh and Abramowicz, 1997, Livio, Ogilvie and Pringle 1998). On the other hand, the torus looks ideal as the source of a dynamo: its large shear rate, the presence of the Balbus-Hawley instability to convert poloidal into toroidal flux, and the possible presence of the anti-floating mechanism inhibiting ballooning of the magnetic field (Kluźniak and Ruderman 1998), all seem to favor the existence of a fast dynamo. It should also be remarked that the configuration of the magnetic field in this

problem is known: in fact, the configuration discussed in Thorne *et al.*, 1986 for black holes, only uses the assumptions of steady-state and axial symmetry, and is thus immediately extended to magnetic fields anchored to the torus. What is really required here is a first order study, of the sort published by Tout and Pringle (1992) on angular momentum removal from young, pre-main-sequence stars via magnetic stresses, and on the associated $\alpha - \omega$ dynamo. Until such studies are made, it will be premature to claim that neutrino annihilations are responsible for the powering of GRBs.

4.2 Progenitors

There is no lack of proposed progenitors, but I will discuss only binary neutron mergers (Narayan, Paczyński and Piran 1992), collapsars (Woosley 1993, Paczyński 1998) and SupraNovae (Vietri and Stella 1998, 1999).

Clearly, NS/NS mergers is the best model on paper: it involves objects which have been detected already, orbital decay induced by gravitational wave emission is shown by observations to work as per the theory, and numerical simulations by Janka's group show that a neutrino-powered outflow in baryon-poor matter can be initiated. The major theoretical uncertainties here concern bursts' durations and energetics: all numerical models produce short bursts (≈ 0.1 s) with modest energetics, $E < 10^{51}$ erg. This is a direct consequence of the mechanism for powering the burst: large, super-Eddington luminosities are carried away by neutrinos, leading to a large mass influx, but only a small fraction, 1 – 3%, can be harnessed for the production of the burst. Furthermore, we cannot invoke large beaming factors in this case: the outflow is only marginally collimated, in agreement with expectations that an accretion disk with inner and outer radii $R_{out}/R_{in} \approx$ a few (for the case at hand) can only produce a beam semi-opening angle of R_{in}/R_{out} . So, perhaps, this model may account for the short bursts, but it should be remembered that nothing of what was discussed above pertains to this subclass: BeppoSAX (and thus all BeppoSAX-triggered observations) can only detect long bursts.

On the other hand, future space missions, whether or not able to locate short bursts, can provide a decisive test of this model, provided they can follow with sufficient sensitivity a given burst for several hours. This model, in fact, is the only one proposed so far according to which some explosions should take place outside galaxies: according to Bloom, Sigurdsson and Pols (1999), about

50% of all bursts will be located more than 8 *kpc* from a galaxy, and 15% in the IGM. This characteristic is testable without recourse to optical observations. In fact, the afterglow begins with a delay (as seen by an outside observer) of $t_d = (r_{ag} - r_{sh})/\gamma^2 c \approx r_{ag}/\gamma^2 c$, which varies greatly depending upon the environment in which the burst takes place:

$$t_d = \begin{cases} 15 \text{ s} & \text{ISM, } n = 1 \text{ cm}^{-3} \\ 5 \text{ min} & \text{galactic halo, } n = 10^{-4} \text{ cm}^{-3} \\ 4 \text{ h} & \text{IGM, } n = 10^{-8} \text{ cm}^{-3} \end{cases} \quad (4)$$

Between the burst proper and the beginning of the power-law-like afterglow, thus a silence of recognizable duration is expected (Vietri 2000).

Collapsars are currently in great vogue as a possible source of GRBs: the large amount of energy available as the core of a supermassive star collapses directly to a black hole is in fact very attractive, even though (again!) the limited efficiency of the reaction $\nu + \bar{\nu} \rightarrow e^- + e^+$ makes most of this energy unavailable. Here too there is some evidence that these objects must exist (Paczynski 1998), and numerical simulations again showing energy preferentially deposited along the hole rotation axis are also available (McFayden and Woosley 1999). Here however, what is truly puzzling is how the outflow can pierce the star's outer layers without loading itself with baryons: we should remember that at most $10^{-6} M_\odot$ can be added to 10^{51} erg : more baryons imply a proportionately slower outflow. The argument is that the dynamical timescale of the outer layers of a massive stars is of order of a few hours, so that, even if the core collapses and pressure support is removed, nothing will happen during the energy release phase: the outflow must pierce its way through. Two processes seem especially dangerous: Rayleigh–Taylor instability of the fluid heated-up by neutrino annihilations as it is weighed upon by the colder, denser outer layers, and Kelvin–Helmholtz instability after the hot fluid has pierced the outer layers and is passing through the hole. It is well-known that the non-linear development of these instabilities leads to mass entrainment, and that the time-scale for the development of these instabilities is very fast. Furthermore, the baryon-free outflow may be ‘poisoned’ by baryons to a deadly extent, even if numerical simulations, with their finite resolution, were to detect nothing of the kind.

The third class of models, SupraNovae, concerns supramassive neutron stars which are stabilized against self-gravity by fast rotation, to such an extent

that they cannot be spun down to $\omega = 0$ because they implode to a black hole. As the star's residual magnetic dipole sheds angular momentum, this is exactly the fate to be expected for the whole star, except for a small equatorial belt, whose later accretion will power the burst. It is easy to show that this implosion must take place in a very baryon-clean environment. The major uncertainties here concern the channels of formation and the existence of this equatorial belt. Two channels of formation have been proposed: direct collapse to a supramassive configuration (Vietri and Stella 1998) and slow mass accretion in a low-mass X-ray binary (Vietri and Stella 1999). Both are possible, though none yet is supported by observations. The existence of the left-over belt has recently been questioned by Shibata, Baumgarte and Shapiro (1999), who however simulated the collapse of neutron stars with intermediate equations of state, which are entirely (or nearly exactly so) contained inside the marginally stable orbit even before collapse: clearly, these must be swallowed whole by the resulting black hole. Soft equations of state are free of this objection, and are thus much more likely to leave behind an equatorial belt. The soft EoSs are especially favored since the neutron stars must survive the r -mode instability, and thus soft EoSs (Weber 1999) would be in any case required. So one might say that the existence of these stars hinges on one uncertainty only, the EoS of nuclear matter. Besides the baryon-clean environments, SupraNovae have another advantage over rival models: only the lowest density regions would be left behind, precisely those with the smallest neutrino losses. The powering of the burst can thus occur through accretion caused by removal of angular momentum by magnetic stresses, without the parallel, unproductive, neutrino generation.

5 Conclusions

It is difficult to end on an upbeat note: we cannot expect in the near future a rate of progress similar to the one we witnessed in the past three years. In particular, it may be expected that the next flurry of excitement will come with the beginning of the SWIFT mission, which promises to collect relevant data (redshifts, galaxy types, location within or without galaxies, absorption or emission features in the optical and in the X-ray) for a few hundred bursts. This data will nail the major characteristics of the environment (at large) in which bursts take place, and we may be able to rule out a few models. On the other

hand, the energy release process, shrouded as it is in optical depths $> 10^{10}$, will remain mysterious, our only hope in this direction being gravitational waves.

Judging by the analogy with radio pulsars, this will correspond to the flattening of the learning curve. Aside from this, we may hope to locate the equivalent of the binary radio pulsar, but, differently from Jo Taylor, we have to be awfully quick in grabbing it.

6 Acknowledgements

Thanks are due to Gabriele Ghisellini, who wisely steered me away from synchrotron emission, and toward the true light of Inverse Compton.

References

1. Baring, M., 1993, Ap.J., 418, 391.
2. Berezhinsky, V.S., Prilutskii, O.F., 1986, Astr.Ap., 175, 309.
3. Bloom, J.S., Sigurdsson, S., Pols, O.R., 1999, MNRAS, 305, 763.
4. Costa, E., *et al.*, 1997, Nature, 387, 783.
5. Covino, S., *et al.*, 1999, Astr.Ap., 348, L1.
6. Fishman, G.J., Meegan, C.A., 1995, ARAA, 33, 415.
7. Frail, D., *et al.*, 1997, Nature, 389, 261.
8. Frail, D., Waxman, E., Kulkarni, S., 2000, Ap.J., 537, 191.
9. Frontera, F., *et al.*, 2000, Ap.J.S., 127, 59.
10. Galama, T., *et al.*, 1998, Ap.J.L., 500, L97.
11. Ghisellini, G., Celotti, A., 1998, Ap.J.L., 511, L93.
12. Ghosh, P., Abramowicz, M.A., 1997, MNRAS, 292, 887.
13. Goodman, J.J., 1997, New Astr., 2, 449.
14. Janka, T., Eberl, T., Ruffert, M., Fryer, C.L., 1999, Ap.J.L., 527, L39.
15. Katz, J., 1994, Ap.J., 422, 248.

16. Klebesadel, R.W., Strong, I.B., Olson, R.A., 1973, Ap.J.L., 182, L85. ,
17. Kluzniak, W., Ruderman, M., 1998, Ap.J.L., 505, L113.
18. Livio, M., Ogilvie, G.I., Pringle, J.E., 1999, Ap.J., 512, 100.
19. Lazzati, D., Campana, S., Ghisellini, G., MNRAS, 1999, 304, L31.
20. Mc Fayden, A., Woosley, S.E., 1999, Ap.J., 524, 262.
21. Mészáros, P. Laguna, P., Rees, M.J., 1993, Ap.J., 415, 181.
22. Mészáros, Rees, M.J., 1993, Ap.J., 405, 278.
23. Mészáros, Rees, M.J., 1997, Ap.J., 476, 232.
24. Narayan, R., Paczynski, B., Piran, T., 1992, Ap.J.L., 395, L83.
25. Paczyński, B., 1986, Ap.J.L., 308, L43.
26. Paczyński, B., 1998, Ap.J.L., 494, L45.
27. Paczyński, B., Xu, G., 1994, Ap.J., 427, 708.
28. Piro, L., *et al.*, 1999, Ap.J.L., 514, L73.
29. Preece, R.D., *et al.*, 2000, Ap.J.S., 126, 19.
30. Rees, M.J., Begelman, M.C., Blandford, R.D., Phinney E.S., 1984, Nature, 295, 17.
31. Rees, M.J., Mészáros, P., 1992, MNRAS, 258, P41.
32. Rhoads, J., 1997, Ap.J.L., 487, L1.
33. Ruderman, R., 1975, Ann. NY. Acad. Sci., 262, 164.
34. Sari, R., Piran, T., 1997, Ap.J., 485, 270.
35. Shibata, M., Baumgarte, T.W., Shapiro, S.L., 1999, Phys.Rev.D, submitted, astro-ph. 9911308.
36. Thorne, K.S., Price, R.H., MacDonald, D.A., 1986, *Black holes: the membrane paradigm*, Yale:New Haven, Yale Univ. Press.

37. Tout, C.A., Pringle, J.E., 1992, MNRAS, 256, 269.
38. Usov, V.V., 1992, Nature, 357, 472.
39. Usov, V.V., Chibisov, G.V. Soviet Astr., 19, 115.
40. van Paradijs, J., *et al.*, 1997, Nature, 386, 686.
41. Vietri, M., 1997a, Ap.J.L., 478, L9.
42. Vietri, M., 1997b, Ap.J.L., 488, L105.
43. Vietri, M., 2000, Astrop. Phys., 14, 211.
44. Vietri, M., Perola, G.C., Piro, L., Stella, L., 1999, MNRAS, 308, P29.
45. Vietri, M., Stella, L., 1998, Ap.J.L., 507, L45.
46. Vietri, M., Stella, L., 1999, Ap.J.L., 527, L43.
47. Waxman, E., 1997, Ap.J.L., 489, L33.
48. Waxman, E., Draine, B.T., 2000, Ap.J., 537, 796.
49. Waxman, E., Frail, D., Kulkarni, D., 1998, Ap.J., 497, 288.
50. Weber, F., 1999, in *Pulsars as astrophysical laboratories for nuclear and particle physics*, Bristol, U.K.; Institute of Physics.
51. Wijers, R.A.M.J., *et al.*, 1999, Ap.J.L., 523, L33.
52. Woosley, S., 1993, Ap.J., 405, 273.
53. Yoshida, A., *et al.*, 1999, Astr.Ap.S., 138, 433.

THE THREAT TO LIFE FROM ETA CARINAE AND GAMMA-RAY BURSTS

Arnon Dar

Department of Physics and Space Research Institute, Technion, Haifa 32000, Israel

A. DeRújula

Theory Division, CERN, Geneva 23, Switzerland

ABSTRACT

Eta Carinae —a large blue variable star in the Carina constellation, more than 100 times as massive and 5 million times as radiant as the Sun— is the most massive and luminous star known in our galaxy [1]. Eta Carinae is rapidly boiling matter off its surface; at any time its core could collapse into a black hole, which may result in a giant supernova and a gamma-ray burst (GRB) [2]. If pointing in our direction, a GRB from Eta Carinae, —only $D \sim 2$ kpc away— would devastate life on Earth. Auspiciously, recent observations indicate that the γ -rays in GRBs are narrowly beamed in cones along the rotational axis of the progenitor star [3,4]. In the case of Eta Carinae the GRBs will not point to us, but will be ravaging to life on planets in our galaxy that happen to lie within the two beaming cones. The mean rate of massive life extinctions by jets from GRBs, per life-supporting planet in galaxies like ours, is once in 100 million years, comparable to the rate observed in our planet [5].

1 Cosmological GRBs

Gamma ray bursts are short-duration flares of MeV γ -rays from outer space that last between a few milliseconds and ~ 1000 s and occur at a rate of about 3 a day [6]. They were discovered serendipitously in 1967 by the Vela satellites launched by the US to monitor the compliance with the Nuclear Test Ban Treaty, banning nuclear explosions in and above the atmosphere. Their exact locations —and consequently their distance and total energy output— were unknown for 30 years, although their isotropy, established by observations with the BATSE instrument on board the Compton Gamma Ray Observatory satellite (CGRO) strongly suggested [7] cosmological distances [8]. Combined with the observed short-time variability of GRBs, such distances imply an enormous energy release from a small volume, if due to spherical explosions. Alternatively, it was argued that if GRB progenitors are so distant, they must be produced by narrow relativistic jets, from the birth of neutron stars or of black holes [9].

The atmosphere is opaque to high energy γ -rays and cosmic ray nuclei, and protects life on Earth from their incoming constant flux. Collisions in the upper atmosphere, however, produce a flux of energetic muons that reach sea level, about 10^{-2} muons $s^{-1} cm^{-2}$. Life on Earth, apparently, has adjusted to the radiation damage from this small flux of atmospheric muons, each depositing through ionization, in biological materials, about $2.4 MeV g^{-1}$. But, if very large fluxes of γ -rays and cosmic ray nuclei suddenly impinge on the atmosphere, they can have a devastating effect on life on Earth. In fact, it has been argued [10] that the highly beamed γ -rays and cosmic rays from GRBs in our galaxy, that happen to point in our direction, can produce lethal fluxes of atmospheric muons at ground level, underground and underwater, destroy the ozone layer and radioactivate the environment, so that GRBs could have caused some of the massive life extinctions on planet Earth in the past 500 My.

Before 1997 the above arguments were mere speculation. But supporting observational evidence accumulated after the significant discoveries of long-lasting GRB X-ray, optical and radio “afterglows”, made possible by the precise and prompt localization of GRBs by the Italian–Dutch satellite BeppoSAX [11]. The GRB and afterglow observations have shown beyond doubt that the “long” duration GRBs (which are the majority) take place in distant galaxies [12], mainly in star formation regions, and are associated with supernova explosions

[13,4]. Their large inferred energies, the properties of their afterglows, their apparent association with supernovae, and their global rate of ~ 1000 per year, imply that GRBs are highly beamed [4].

Because of the limited sizes of the satellite-borne detectors, GRBs have been observed mostly in the sub-MeV energy region, where the photon number flux, decreasing with increasing energy, is large enough. However, for a few very bright GRBs the EGRET instrument on board the CGRO detected γ -rays of up to GeV energies [14]. Moreover, four large ground-based γ -ray detectors, the Tibet air shower array, the HEGRA-AIROBICC Čerenkov array, the Milagro water-Čerenkov detector, and GRANDE, have reported possible detections of TeV γ -rays in directional and temporal coincidence with some GRBs detected by BATSE. In every case, the estimated total energy in TeV photons was about 2 orders of magnitude larger than the energy in sub-MeV photons measured by BATSE. In particular, GRANDE [15] and MILAGRITO [16] have reported the detection of unexpectedly large fluxes of muons coincident in time and direction with GRBs. These muons are allegedly produced by the interactions in the upper atmosphere of γ -rays from the GRB with energies well above 100 GeV. These observations, if confirmed, would imply that GRBs are more lethal than they were previously thought to be. Since TeV photons are absorbed in the intergalactic infrared (IR) background by pair production, only relatively close-by GRBs (for which this absorption is insignificant) can be observed at TeV energies. This may explain why only a small fraction of the BATSE-detected GRBs in the fields of view of the various ground-based detectors were claimed to have been seen at TeV energies.

In Table I we list the measured redshift and fluence F_γ (in units of 10^{-5} erg cm^{-2}) in the BATSE energy band, 40–2000 keV, for all GRBs with known redshift z . We also list their inferred luminosity distance D_L (in units of Gpc) and their total energy output, $E_\gamma = 4\pi D_L^2 F_\gamma / (1+z)$ (in units of 10^{53} erg), assuming isotropic emission and a critical Universe with a Hubble constant $H_0 = 65 \text{ km s}^{-1} \text{ Mpc}^{-1}$, fractional matter density $\Omega_M = 0.3$ and vacuum energy density $\Omega_\Lambda = 0.7$.

2 GRB from the Death of Eta Carinae

Should the violent end of Eta Carinae, the most massive star known in our galaxy and only $D = 2 \text{ kpc}$ away, emit in our direction a GRB similar to that

of the most energetic GRB in Table I (GRB 990123), the atmosphere of Earth facing the star would be subject to a total energy deposition:

$$\frac{E_\gamma}{4\pi D_L^2} \approx 4 \times 10^9 \text{ erg cm}^{-2} \quad (1)$$

within seconds. This energy release is akin to that of the simultaneous explosions in the upper atmosphere of one-kiloton of TNT per km², over the whole hemisphere facing Eta Carinae. This would destroy the ozone layer, create enormous shocks going down in the atmosphere, lit up huge fires and provoke giant global storms.

If the energy of GRBs in TeV γ -rays, as indicated by various experiments [15,16], is ~ 100 times larger than in the sub-MeV domain, the energy deposition of Eq. 1 would be correspondingly larger. Moreover, the interactions of the TeV γ -rays in the upper atmosphere would produce a lethal dose of highly penetrating muons, destroying life on the surface, underground and underwater. Indeed, a high energy γ -ray impinging on the atmosphere at a large zenith angle θ produces $\sim 0.23 \cos \theta [\epsilon_\gamma]^{1.17}$ muons at ground level [15], where ϵ_γ is the γ -ray energy in TeV. Hence, the total muon fluence at ground level expected from a GRB from the supernova death of Eta Carinae is $\sim 5 \times 10^{10} \text{ cm}^{-2}$ (the roughly linear dependence on the γ -ray energy makes this result sensitive only to the total deposited energy). The energy deposition by these high-energy muons in biological materials is $\sim 2.5 \times 10^5 \text{ erg g}^{-1}$, which is about ten times the lethal dose for human beings: the whole-body dose from penetrating ionizing radiation resulting in 50% mortality in 30 days.

All of the above, which would be devastating for life on Earth, would only happen if the γ -rays from Eta Carinae's supernova point in our direction. But would this GRB point to us? There are at least three known superheavy stars in our galaxy with a lifetime shorter than $\sim 1 \text{ My}$ and expected to end in a giant supernova, implying that the galactic rate of giant supernovae is $\geq 3 \times 10^{-6} \text{ y}^{-1}$. The rate of massive life extinctions is $\sim 10^{-8} \text{ y}^{-1}$. Thus, if all galactic giant supernovae produced deadly GRBs, their γ -rays must be funnelled in a cone of opening angle $\theta_b \leq 5^\circ$, for two opposite GRBs per giant supernova. The chance probability for such cones to point in our direction is only 3×10^{-3} . But the expected direction for a jetted GRB [4] is the progenitor's polar axis, which for Eta Carinae points $57^\circ \pm 10^\circ$ away from our direction, judging from the radial velocities, proper motions and projected shape of its equatorial disk

of debris [1]. This reduces considerably the chance that the GRBs from Eta Carinae point to our planet. Moreover, the properties of GRB afterglows and their association with Type Ib/Ic supernovae imply that GRBs are beamed into much narrower cones, of 1 mrad typical opening angle! [4]. This reduces to a negligible level the threat to terrestrial life from Eta Carinae.

3 Mass Extinctions by Galactic GRBs

Could Galactic GRBs beamed in our direction have caused some of the massive life extinctions in the history of Earth? The average energy output of a GRB is 5 times smaller than that of GRB 990123, as can be seen in Table I. The average distance of galactic GRBs from Earth, assuming they have the same spatial distribution as supernova remnants, is ~ 8 kpc. Gamma rays alone from such “typical” GRBs can barely cause major mass extinctions, since the frequency of such GRBs is too small to explain a mean rate of mass extinctions of once in ~ 100 My, observed in the geological records. However, if GRBs are produced in supernova explosions by highly relativistic jets of “cannonballs”, as suggested by the striking success of the Cannonball Model of GRBs in explaining their afterglows [4], the jetted cannonballs also produce highly beamed cosmic rays (CRs) by ionizing, sweeping up and accelerating the particles of the interstellar medium. Such CRs from galactic GRBs are much more devastating than their γ -rays. Let v be the speed of the CB and $\Gamma \equiv 1/\sqrt{1 - (v/c)^2} \gg 1$ be its Lorentz factor. The bulk of the swept up ISM particles entering the CB with energy $\Gamma m c^2$ in its rest frame are deflected by the CB’s tangled magnetic fields, and are emitted isotropically in that frame. In the galactic rest frame their energy is Lorentz-boosted to an average energy $m c^2 \Gamma^2$ and they are beamed into a cone of opening angle $\theta \sim 1/\Gamma$. Their energy distribution is related to the CBs’ deceleration by energy-momentum conservation, which yields $dN_{\text{CR}}/d\Gamma \approx N_{\text{CB}}/\Gamma^2$, where N_{CB} is the baryonic number of the CBs [4]. The afterglows of the GRBs listed in Table I are very well fitted [4] with initial Lorentz factors $\Gamma_i \simeq 10^3$ and total baryonic number $N_{\text{jet}} \sim 6 \times 10^{50}$, comparable to that of the Earth. Thus, the energy fluence of CRs within their beaming cone of opening angle $\theta \leq \Gamma_i$, from a galactic GRB at a distance $d \sim 8$ kpc, is:

$$F \simeq \frac{E_{\text{jet}} \Gamma_i^2}{3\pi d^2} \simeq 1.5 \times 10^{12} \text{ erg cm}^{-2}. \quad (2)$$

Most of this fluence is spread over less than $\Delta t \sim 2$ days, the typical CB deceleration time [4] from $\Gamma = \Gamma_i$ to $\Gamma = \Gamma_i/2$. It is carried by CRs with energies between $E = 2 m_p c^2 \Gamma_i^2 \sim 2 \times 10^3$ TeV and $E = 0.4 m_p c^2 \Gamma_i^2/4 \sim 4 \times 10^2$ TeV.

The ambient interstellar gas is transparent to the CR beam because the Coulomb and hadronic cross sections are rather small with respect to typical galactic column densities. Although the galactic magnetic field, $B \sim 5 \times 10^{-6}$ Gauss, results in a Larmor radius $r_L = \beta E_p / c q B \leq 10^{18}$ cm $\ll 8$ kpc for single protons with $E_p \leq 10^{15}$ eV, it does not deflect and disperse the CR beams from galactic GRBs. This is because of the high collimation of the CR beam which, even after travelling for a typical galactic distance —e.g. $d \sim 8$ kpc, our distance from the Galaxy's centre— has a very large energy and pressure within an angle $\theta \leq 1/\Gamma_i$ from its direction of motion: $E_{\text{CR}} \sim E_{\text{jet}}/3 \sim 3 \times 10^{51}$ erg and $P_{\text{CR}} \sim E_{\text{jet}}/(3 \pi d^2 c \Delta t) \sim 3 \times 10^{-4}$ erg cm $^{-3}$, respectively. These figures are much larger than the total magnetic energy of the swept-up galactic magnetic field inside the cone, $d^3 B^2/24 \Gamma_i^2 \sim 1.5 \times 10^{49}$ erg and the galactic magnetic pressure $B^2/8 \pi \sim 10^{-12}$ erg cm $^{-3}$. Thus, the CR beam sweeps away the magnetic field along its way and follows a straight ballistic trajectory through the interstellar medium. (The corresponding argument, when concerning the distant cosmological GRBs, reaches the opposite conclusion: no CRs accompany the GRB in all of its voyage here.)

The beam of multi-TeV cosmic rays accompanying a galactic GRB is deadly for life on Earth-like planets. The total number of high energy muons ($E_\mu \geq 25$ GeV) in the atmospheric showers produced by a cosmic ray proton with energy $E_p \sim 10^2$ to 10^3 TeV is $N_\mu(E > 25 \text{ GeV}) \sim 9.14 [E_p/\text{TeV}]^{0.757}/\cos\theta$ [17], yielding a muon fluence at ground level:

$$F_\mu(E > 25 \text{ GeV}) \simeq 1.7 \times 10^{12} \text{ cm}^{-2}. \quad (3)$$

Thus, the energy deposition rate at ground level in biological materials, due to exposure to atmospheric muons produced by an average GRB near the centre of the Galaxy, is 4.2×10^{12} MeV g $^{-1}$. This is approximately 270 times the lethal dose for human beings. The lethal dosages for other vertebrates and insects can be a few times or as much as a factor 20 larger, respectively. Hence, CRs from galactic GRBs can produce a lethal dose of atmospheric muons for most animal species on Earth. Because of the large range of muons ($\sim 4 [E_\mu/\text{GeV}]$ m in water), their flux is lethal, even hundreds of metres underwater and underground, for CRs arriving from well above the horizon. Thus, unlike other suggested

extraterrestrial extinction mechanisms, the CRs of galactic GRBs can also explain massive extinctions deep underwater and underground. Although half of the planet is in the shade of the CR beam, its rotation exposes a larger fraction of its surface to the CRs, whose arrival time is spread over ~ 2 days. Additional effects increase the lethality of the CRs over the whole planet. They include:

- (a) Environmental pollution by radioactive nuclei, produced by spallation of atmospheric and surface nuclei by the secondary particles of the CR-induced showers.
- (b) Depletion of stratospheric ozone, which reacts with the nitric oxide generated by the CR-produced electrons (massive destruction of stratospheric ozone has been observed during large solar flares, which generate energetic protons).
- (c) Extensive damage to the food chain by radioactive pollution and massive extinction of vegetation by ionizing radiation (the lethal radiation dosages for trees and plants are slightly higher than those for animals, but still less than the flux given by Eq. 3 for all but the most resilient species).

4 The Rate of Galactic GRBs

Are the geological records of mass extinctions consistent with the effects induced by cosmic rays from GRBs? Good quality geological records, which extend up to ~ 500 My ago, indicate that the exponential diversification of marine and continental life on Earth over that period was interrupted by many extinctions [5], with the major ones —exterminating more than 50% of the species on land and sea— occurring on average every 100 My. The five greatest events were those of the final Ordovician period (some 435 My ago), the late Devonian (357 My ago), the final Permian (251 My ago), the late Triassic (198 My ago) and the final Cretaceous (65 My ago). The observed rate of GRBs is $\sim 10^3 \text{ y}^{-1}$. The sky density of galaxies brighter than magnitude 25 (the observed mean magnitude of the host galaxies of the GRBs with known redshifts) in the Hubble telescope deep field is $\sim 2 \times 10^5$ per square degree [18]. Thus, the rate of observed GRBs, per galaxy with luminosity similar to that of the Milky Way, is $R \sim 1.2 \times 10^{-7} \text{ y}^{-1}$. To translate this result into the number of GRBs born in our own galaxy, pointing to us, and occurring at (cosmologically) recent times, one must take into account that the GRB rate is proportional to the star formation rate, which increases with redshift like $(1+z)^3$. For GRBs with known redshift (see Table I) one finds $\langle 1+z \rangle \sim 2.1$. In a flat Universe

(like ours) the probability of a GRB to point to us within a certain angle is independent of distance. Therefore, the mean rate of GRBs pointing to us and taking place in our galaxy is roughly $R/(1+z)^3 \sim 1.3 \times 10^{-8} \text{ y}^{-1}$, or once every $\sim 70 \text{ My}$. If most of these GRBs take place not much farther away than the distance to the galactic centre, their effect is lethal, and their rate is consistent with the rate of the major mass extinctions on our planet in the past 500 My.

5 Mass Extinctions and Coincident Catastrophes

The geological records also indicate that two of the major mass extinctions were correlated in time with impacts of large meteorites or comets, with gigantic volcanic eruptions, with huge sea regressions and with drastic changes in global climate. A large meteoritic impact was invoked [19] in order to explain the iridium anomaly and the mass extinction that killed the dinosaurs and claimed 47% of existing genera at the Cretaceous-Tertiary (K/T) boundary, 65 My ago. Indeed, a 180 km wide crater was later discovered, buried under 1 km of Cenozoic sediments, dated back 65 My ago and apparently created by the impact of a $\sim 10 \text{ km}$ diameter meteorite or comet near Chicxulub, in the Yucatan [20]. The huge Deccan basalt floods in India also occurred around the K/T boundary 65 My ago [21]. The Permian/Triassic (P/T) extinction, which killed between 80% and 95% of the species, is the largest known in the history of life [22]; it occurred 251 My ago, around the time of the gigantic Siberian basalt flood. Recently, evidence was found [23] for a large cometary impact at that time.

The orbits of comets indicate that they reside in a spherical cloud at the outer reaches of the solar system –the Oort Cloud– with a typical radius of $R_O \sim 50000 \text{ AU}$. The statistics imply that it may contain as many as 10^{12} comets with a total mass perhaps larger than that of Jupiter. The large value of R_O implies that the comets have very small binding energies and mean velocities of $v \sim 100 \text{ m s}^{-1}$. Small gravitational perturbations due to neighbouring stars are believed to disturb their orbits, unbind some of them, and put others into orbits that cross the inner solar system. The passage of the solar system through the spiral arms of the Galaxy where the density of stars is higher, could also have caused such perturbations and consequently the bombardment of Earth with a meteorite barrage of comets over an extended

period longer than the free fall time from the Oort cloud to the Sun:

$$t_{\text{fall}} = \pi \left[\frac{R_{\text{O}}^3}{8GM_{\odot}} \right]^{1/2} \simeq 1.7 \text{ My}. \quad (4)$$

The impact of comets and meteorites from the Oort cloud could have triggered the huge volcanic eruptions that created the observed basalt floods, timed — within 1 to 2 My— around the K/T and P/T boundaries. Global climatic changes and sea regression followed, presumably from the injection of large quantities of light-blocking materials into the atmosphere, from the cometary impacts and the volcanic eruptions. In both the gigantic Deccan and Siberian basalt floods $\sim 2 \times 10^6 \text{ km}^3$ of lava were ejected. This is orders of magnitude larger than in any other known eruption, making it unlikely that the other major mass extinctions, which are of a similar magnitude, were produced by volcanic eruptions. The volcanic-quiet and impact-free extinctions could have been caused by GRBs. Moreover, passage of the GRB jet through the Oort cloud after sweeping up the interstellar matter on its way could also have generated perturbations, sending some comets into a collision course with Earth, perhaps explaining also the geologically active K/T and P/T extinctions.

6 An Answer to Fermi’s Question: Where Are They?

The observation of planets orbiting nearby stars has become almost routine, but current techniques are insufficient to detect planets with masses comparable to the Earth’s. Future space-based observatories to detect Earth-like planets are being planned. Terrestrial planets orbiting in the habitable neighbourhood of stars, where planetary surface conditions are compatible with the presence of liquid water, might have global environments similar to ours, and harbour life. Our solar system is billions of years younger than most of the stars in the Milky Way. Life on extrasolar planets could have preceded life on Earth by billions of years, allowing for civilizations much more advanced than ours. Thus Fermi’s famous question “where are they?”, i.e. why did they not visit us or send signals to us? An answer is provided by GRB-induced mass extinctions: even if advanced civilizations are not self-destructive, GRBs can exterminate the most evolved species on any given planet or interstellar vehicle at a mean rate of once every 100 My. Consequently, there may be no nearby aliens having evolved long enough to be capable of communicating with us, or pay us a visit.

Acknowledgement:

The authors thank LeV Okun for useful comments. The partial support by the Helen Asher Fund for Space Research and by the Technion VPR Fund - Steiner Fund for the promotion of research is gratefully acknowledged.

Table I - GRBs of known redshift

GRB	z	D_L	F_γ	E_γ
970228	0.695	4.55	1.1	0.22
970508	0.835	5.70	0.32	0.07
970828	0.957	6.74	9.6	2.06
971214	3.418	32.0	0.94	2.11
980425	.0085	.039	0.44	8.1E-6
980613	1.096	7.98	0.17	0.61
980703	0.966	6.82	2.26	1.05
990123	1.600	12.7	26.8	19.80
990510	1.619	12.9	6.55	5.00
990712	0.434	2.55	6.5	0.53
991208	0.70	4.64	10.0	1.51
991216	1.020	7.30	19.4	5.35
000131	4.500	44.4	4.2	11.60
000301c	2.040	17.2	0.41	0.46
000418	1.119	8.18	2.0	0.82
000911	1.06	7.66	2.0	0.68
000926	2.066	17.4	2.20	10.54
010222	1.474	11.5	12.0	7.80

Redshift z . Luminosity distance, D_L , in Gpc. Fluence measured by BATSE, F_γ , in 10^{-5} erg cm^{-2} units. Deduced spherical energy, E_γ , in 10^{53} erg units.

References

1. K. Davidson, and R.M. Humphrey, *Ann. Rev. Astron. Astrophys.* **35**, 1 (1997).
2. S.E. Woosley, *Astrophys. Jour.* **405**, 273 (1993).
3. A. Dar, *Astrophys. Jour.* **500** L93 (1998).

4. S. Dado, A. Dar, and A. De Rújula, astro-ph/0107367 (2001) and references therein.
5. M.J. Benton, *Science*, **268**, 52 (1995).
6. C.A. Meegan, and G.J. Fishman, *Ann. Rev. Astron. Astrophys.* **33**, 415 (1995).
7. C.A. Meegan, *et al*, *Nature* **355**, 143 (1992).
8. V.V. Usov, and G.V. Chibisov, *Astronomicheskii Zhurnal* **52**, No. 1, 192 (1975).
9. N. Shaviv and A. Dar, *Astrophys. Jour.* **447**, 863 (1995).
10. A. Dar, A. Laor and N. Shaviv, *Phys. Rev. Lett.* **80**, 5813 (1998).
11. E. Costa *et al*, *Nature* **387** 783 (1997).
12. M.R. Metzger, *et al*, *Nature*, **387** 878 (1997).
13. T.J. Galama, *Nature* **395**, 670 (1998).
14. K. Hurley, *Nature* **372**, 652 (1994).
15. T.F. Lin *et al*, **ICRC26 Vol. 4**, 24 (1999).
16. R. Atkins et al, *Astrophys. Jour.* **533**, L119 (2000).
17. M. Drees, F. Halzen and K. Hikasa, *Phys. Rev.* **D39**, 1310 (1989).
18. S. Casertano et al, *Astron. Jour.* **120**, 2747 (2000).
19. L.W. Alvarez *et al*, *Science* **208**, 1095 (1980).
20. J. Morgan, et al. *Nature* **390** 472 (1997).
21. C.B. Officer *et al*, *Nature* **326**, 143 (1987).
22. D.H. Erwin, *Scientific American* **275**, July 56 (1996).
23. L. Becker et al, *Science* **291** 1530-1533 (2001).

GAMMA-RAY SATELLITES AND GAMMA-RAY BURSTS

Francesco Longo

Univ. di Ferrara and INFN, sezione di Ferrara, Italy
on the behalf of the AGILE collaboration

ABSTRACT

The EGRET experiment on board the Compton Gamma-Ray Observatory has extended the observation of GRB emission to GeV gamma-rays, however only for a few events, due to its limited field of view and effective area. The high energy emission is related to several issues regarding GRB physics, such as relativistic motion, baryon loading, internal versus external shocks and emission mechanism. GRBs may be able also to probe some Quantum Gravity effects. Some of these implications as well as the strategy of the new gamma-ray satellites AGILE and GLAST will be outlined in this paper.

1 Introduction

Gamma-ray bursts (GRB) are the most intense and most distant known sources of high-energy gamma rays. After more than thirty years of study on GRB, the

detection of afterglow at lower frequencies (X-rays, optical, infrared, radio) has demonstrated that some GRB sources are cosmological, as previously supposed thanks to the measure of GRB spatial isotropy and LogN-LogS distribution by BATSE 1, 2, 3).

Popular models for the GRB central engine are generally based on black-hole accretion disk scenarios and the endpoints of stellar evolution of massive stars. In all the models the total energy released and the volume of the emitting source require a relativistic fireball. The observations at different wavelengths generally confirm this model 4).

In the fireball models, higher energy pulses are known to be narrower and to peak earlier than at higher energies 5, 6). However the high energy spectral evolution is very poorly known, owing to precedent instruments long deadtime and relatively small effective area.

2 EGRET observation of GRB

Prior to the launch of the Compton Gamma-Ray Observatory, the spectra of GRBs were known to extend beyond 10 MeV in several bursts, ruling out thermal emission. EGRET has extended the maximum energy emission from GRBs to 20 GeV and detected longer duration emission than measured at MeV energies 7).

EGRET had two modes in 2 overlapping energy ranges to observe GRBs. The higher energies (above 30 MeV) were recorded in the standard mode. Individual gamma rays were detected with their time, direction and energy measured. Due to the small number of gamma-rays detected in a GRB the spectral information were very uncertain. Also only a lower limit on the flux could be determined because of the EGRET dead time of ~ 100 ms per trigger. Four bright GRB were detected by EGRET in this standard mode. The average spectrum is hard with a photon spectral index of 1.95 ± 0.25 . Because these are the brightest BATSE bursts, possibly all GRBs have high energy emission. In one burst, GRB 940217 EGRET detected high energy emission persisting for ~ 5000 s beyond main burst activity at hard-X ray range 9).

EGRET's second mode of observing GRBs uses the Total Absorption Shower Counter (TASC) that could measure energy spectra in the energy range 1 - 100 MeV. Because of large background individual events were not recorded. The spectra of ~ 30 bursts was observed, extending in some cases up to ~ 100

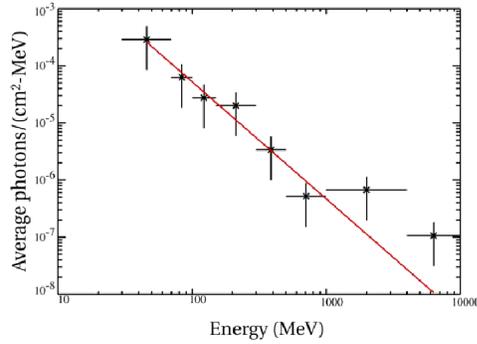


Figure 1: *Average Spectrum of GRB detected by EGRET* ⁷⁾.

MeV. None of these spectra showed an high energy break ⁷⁾.

3 High energy emission and GRB models

The higher energy gamma-rays could constrain the bulk Lorentz factor of the relativistic fireball. Pair production will attenuate high energy gamma flux unless the relativistic corrections are large enough that in the rest frame of the fireball the γ -rays energy is below the threshold of pair production with lower energy photons ⁸⁾.

The burst fireball carries both electrons and protons. Very short duration GeV emission could be produced by fireball with the smallest baryon loading ¹⁰⁾. Clean fireballs decelerate more rapidly producing higher energy gamma-rays while the bulk Lorentz Factor is still high.

Shocks are an efficient method for accelerating particles to high energies. The gamma-ray emission is believed to be due to internal shocks producing the variable light curve at hard-X ray and possibly at gamma-ray energies. External shocks, the probable mechanism for the afterglow emission at lower frequencies, are due to the interactions of the fireball with the external medium whose inhomogeneities will produce variable gamma-ray emission at later stages. Measurement of spectral evolution at high energy will help to distinguish between the two emissions.

The BATSE spectra are well fit by a low-energy power law breaking to

a steeper high energy power law. These spectra show only a single peak generally interpreted as a synchrotron peak as seen in other high energy sources like AGNs. Higher energy observations could confirm the existence of the second bump and constrain the emission mechanisms like pion decay or inverse compton scattering.

GRB may be able to probe the scale of quantum gravity at the Planck Length scale ¹¹⁾. If the space-time is granular at that scale, the light should produce an energy-dependent dispersion of the order of $10 \text{ ms GeV}^{-1} \text{ Gpc}^{-1}$. GRBs could have temporal structures comparable to this timescales at energies greater than 10 MeV.

4 GRB observations with AGILE and GLAST

AGILE is an ASI Small Scientific Mission dedicated to high energy astrophysics ¹²⁾. The AGILE scientific instrument is made of three integrated detectors with broad-band detection and imaging capabilities ^{12, 13)}. The AGILE Gamma-Ray Imaging Detector (GRID), sensitive in the range 30 MeV- 50 GeV, consists of a Silicon-Tungsten Tracker, a Cesium Iodide Mini-Calorimeter, an Anticoincidence system made of segmented plastic scintillators, fast readout and data-handling electronics. The Super-AGILE detector will provide detection and imaging capabilities in the hard X-ray range (10-40 keV). The CsI Mini-Calorimeter will also detect and collect events independently from the GRID in case of impulsive transients in the energy range ~ 0.3 -100MeV. The GRID detection rate of GRBs is expected to be a factor ~ 5 larger than of EGRET. The small GRID deadtime allows a better of the initial phase of GRB pulses. AGILE is expected to be highly efficient in detecting above 10 GeV because of limited backscattering. Super-AGILE will be able to locate GRB within a few arcminutes. Special emphasis will be given to fast timing ¹²⁾.

GLAST is an international space mission that will study high-energy astrophysics in the energy range 10 keV - 300 GeV. GLAST will have an imaging gamma-ray telescope, the Large Area Telescope (LAT) whose effective area, angular resolution, field of view and dead time will provide unique capabilities to study transient phenomena in the energy range 20 MeV - 300 GeV. The Glast Burst Monitor (GBM) will have a field of view several time larger than the LAT and will provide spectral coverage of GRB that extends from the lower

limit of the LAT down to a few keV ¹⁴).

5 GRB simulation studies

To study the capabilities of these two instruments a detailed simulation study is being performed. Regarding the AGILE satellite, a comparison between two different Montecarlo programs sensitive in different but overlapping energy ranges is in progress. With this study a detailed analysis of the GRB trigger logic, dependent both on the SuperAgile and the MCAL signal will be performed. The adopted strategy is to generate a light curve from available data and phenomenological description.

For the GLAST study, the strategy is somehow different. The Geant4 proposed simulation toolkit is adequate to simulate gamma-ray interaction overall the energy range of the two detectors. The GRB event generator that will be adopted is based on the model by Omodei ¹⁵) that generates the light curve of a GRB together with its spectral evolution including also the highest energies emission. With this program a study of the different GRB detection strategies and capabilities of the GLAST satellite is in progress.

References

1. Fishman G.J. and Meegan C.A. (1995), *ARA&A*, 33, 415.
2. Costa E. et al. (1997), *Nature* 387, 783
3. van Paradijs J., Kouveliotou C. and Wijers R.A.M.J. (2000), *ARA&A*, 38, 379
4. Piran T. (1999), *Physics Reports* 314, 575
5. Fenimore E.E. et al. (1995), *ApJ* 448, L101
6. Norris J.P. et al. (1996), *ApJ* 459, 393
7. Dingus B.L. (2001), *High Energy Gamma-Ray Astronomy*, AIP Conference Proceedings 558, 383
8. Baring M.G. and Harding A.K. (1997), *ApJ* 491, 663
9. Hurley K. et al. (1994), *Nature* 372, 652

10. Dermer C.D., Chiang J. and Böttcher M. (1999), ApJ 513, 656
11. Amelino Camelia G. et al. (1998)
12. Tavani M. et al., Proceedings of Gamma2001 Conference, AIP 587, 729
13. Barbiellini G. et al., Proceedings of Gamma2001 Conference, AIP Conference Proceedings 587, 774
14. GLAST science brochure <http://glast.gsfc.nasa.gov>
15. Omodei N. (2002), these Proceedings

Emission Model and GRB Simulations

Nicola Omodei
University of Siena and INFN Pisa

ABSTRACT

We present a simple emission model based on the fireball scenario of Gamma Ray Burst sources. The goal of this work is to obtain a complete chain for simulating the response to GRBs of the *Large Area Telescope* (LAT) detector of GLAST. In this paper we describe the emission model that we adopt to obtain the spectrum and the light curves of the GRB and we spend some time describing the procedure to extract a photons from the simulated flux.

1 Introduction

The observations collected by the Compton Gamma Ray Observatory (CGRO) have provided a great improvement in the GRB knowledge. The typical signature of a GRB is the rapid variability in the first stage of emission as has been observed by BATSE in the soft gamma ray energy range. This prompt emission

can have a typical time scale of the order of few milliseconds ¹⁾, depending on the energy band ³⁾. There are no indications about a similar variability in the high energy range: this data lack is basically due to the experimental limit of the past experiments (EGRET deadtime was greater than $100ms$). With the planned gamma observatory GLAST we will be able to investigate the variability in the signal up to few $100\mu s$, and a great breakthrough will be done in the GRB knowledge. The direction of our studies is to develop a simulation model that includes the emission of radiation at very high energies maintaining the interesting features of the rapid time variability.

2 Outline of the emission model

The highly variable temporal structure observed in most GRBs is thought to be one of the typical signature of the presence of internal shocks ⁴⁾. In the fireball scenario ²⁾ the kinetic energy of shells ejected from a compact source (the central engine) is converted into the observed radiation by shocks. The idea of the internal shocks scenario is that the faster shells can reach the slower ones producing shocks ^{8, 9, 10)}. During each one of these shocks the electrons are efficiently accelerated by the shock wave produced in the inner part of the shell. If the number of shells emitted in the surrounding medium by the central engine is high enough they will bunch together, producing the rapid variability observed in the light curves.

In our model the dimensions of the source (typically $10^6 - 10^8$ cm) are related with the variability at short time scale ($0.1 - 10ms$). The shells are emitted from the engine with a random distribution of Lorentz Factors ($10 < \Gamma < 10^3$). The energy injected in the surrounding medium by the engine for each shell is the ratio between the total energy available during the burst E_{tot} and the number of shells (N_{shells}) emitted. The typical numbers are $E_{tot} \approx 10^{52} ergs$, $100 < N_{shell} < 1000$. The initial separation of the shells, as well as their initial thickness, is of same order of the dimension of the central engine. All these parameter can be set as initial conditions.

The most likely radiation process in GRBs is synchrotron emission ^{5, 6, 7)}. The observed low energy spectra provide an indication that this is indeed the case. The parameters that determine synchrotron emission are the magnetic field strength, B_{eq} , and the electrons' energy distribution. It is difficult to estimate the value of the magnetic field and the amount of energy available for

the electrons from first principles. Instead we define two dimensionless parameters, ϵ_B and ϵ_e , that incorporate our ignorance and uncertainties^{7, 2)}. They correspond to the fractions of total thermal energy that goes into magnetic field energy density and into random motions of the electrons. The accelerated electrons, interacting with the magnetic field, emit by synchrotron radiation with a typical observed energy of:

$$(E_{syn})_{obs} = (\hbar e B / m_e c) \gamma_e^2 \Gamma_f \quad (1)$$

Where γ_e is the Lorentz factor of the accelerated electrons relatives to the shell. According with the Fermi's mechanism the electrons energy distribution that comes out from a shock is a power law of index p ($N_e(E) \propto E^{-p}$), is then possible to calculate the spectrum of a population of such electrons that emits by synchrotron radiation. The calculation is performed by taking into account that the total power radiated by synchrotron in the energy interval between E and $E + dE$ is given by:

$$P(E, E_{syn}, B_{eq}) \approx \begin{cases} h B_{eq} (E/E_{syn})^{1/3} & E < E_{syn} \\ h B_{eq} (E/E_{syn})^{1/2} \exp(1 - E/E_{syn}) & E > E_{syn} \end{cases} \quad (2)$$

Where E_{syn} is given by equation 1. Than we calculate the total flux radiated from the whole population of electrons by integrating over their energy distribution.

$$P_{tot}(E) = \int N_e(E_{syn}) P(E, E_{syn}, B_{eq}) dE_{syn} \quad (3)$$

The natural extension of the spectrum obtained to higher energies is the Inverse Compton mechanism in which the photons, produced by synchrotron radiation, could up-scatter against the high energy electrons producing high energy photons. If $\gamma_e m_e c^2 (\gg E_{syn})$ is the energy of the electron, the energy of the outgoing photon is shifted to higher energies by a quantity equal to γ_e^2 . Figure 1 shows the typical flux obtained from the simulation. In figure 2 four different Light Curves are plotted as a function of time. The curves represent the temporal profiles of the flux in four different energy bands. The curves correspond to the different BATSE's channels and to the energy band of the *LAT* detector on *GLAST* satellite (see the label of figure 2 for details). The model we have developed seems to be able to reproduce the temporal

behaviour observed in the BATSE triggers. The pulse duration depends on the energy band we are considering (i.e. harder spikes have short duration, in agreement with observations ³⁾). Millisecond variability is found also in the highest energy band allowing us to investigate the GLAST capability in detecting rapidly variable signals.

3 The Event Generator

As we have already mentioned the main signatures of the GRB radiation are the rapid temporal variability and the spectral evolution with time. The aim of this work is to develop a tool for generating a photons list to be used as "event generator" for the *GEANT4* simulator of the LAT detector. We refer to the paper by Francesco Longo (these proceedings) for the description of how the simulation of the detector is performed and how the photons list can be used as input of the simulation tools.

Figure 3 is a simplified chart to explain the logical structure of the simulation chain. The source simulation -*the GRB Emission Model* box- calculates the spectral and temporal evolution of a GRB and writes a file that is used as input by the detector simulation. Our code extracts photons from the flux in each time interval; in each bin, photons with different energies are extracted until the amount of energy radiated is reached. The goodness of this procedure can be controlled by filling an histogram containing the number of photons extracted in each energy bin. In figure 4 is shown the histogram overimposed to the calculated amount of photons. The efficiency of this algorithm is evaluated calculating the ratio between the total energy radiated by the source (integrating the fluence in time) and the sum of all energies of all photons. It is always greater than 95%, depending on the brightness of the burst.

4 Conclusions

The idea that we are following is to build up a complete simulation framework in which photons reach the detector carrying the same signature of a GRB signal. In the case of the GRB sources the temporal and the spectral variability observed for energies up to few *MeV* has been extended to higher energies in order to study the response of GLAST observatory. We believe that we could extend the simulation framework to other sources, in particular to the sources

that show a typical temporal variability in their emission (Active Galactic Nuclei, Pulsars, Repeaters, micro Quasars).

A further implementation of the emission model will include the high energy emission from the hadronic component of the fireball (i.e. the π^0 decay), and a more precise description of the interaction between photons and pairs. Two photons annihilation into pairs and pair annihilation into photons are some examples. We are also planning to provide different scenarios to simulate different signatures in the GRB signal. For example the external shock model proposed by Dermer¹¹⁾ provide a natural high energy emission caused by a relativistic reverse shock and a typical temporal behaviour of the spectrum. GLAST will be able to detect the high energy emission from GRBs discriminating theoretical model starting from their predictions at high energy and will fix important constraints on the phenomenology of Gamma Ray Burst sources.

References

1. K.C. Walker *et al*, Ap. J. **537**, 264 (2000).
2. T. Piran, Physics Reports **314**, 575 (1999).
3. E.E. Fenimore *et al*, Ap. J. **448**, L101 (1995).
4. E. Ramirez-Ruiz & E.E. Fenimore, Ap. J. **539**,712 (2000).
5. P. Mészáros *et al*, Ap. J., **415**, 181 (1993).
6. J.I. Katz, Ap. J., **422**, 248 (1994)
7. R. Sari *et al*, Ap.J., **473**, 204 (1996).
8. M. Spada *et al*, Ap. J. **537**, 824S (2000).
9. A. Panaitescu *et al*, Ap. J. **522L**, 105P (1999).
10. D. Guetta *et al*, Ap. J. **557**, 399 (2001).
11. C.D. Dermer, Gamma-ray Bursts 5th Huntsville Sym., AIP Conference Series **526**, 431 (2000).

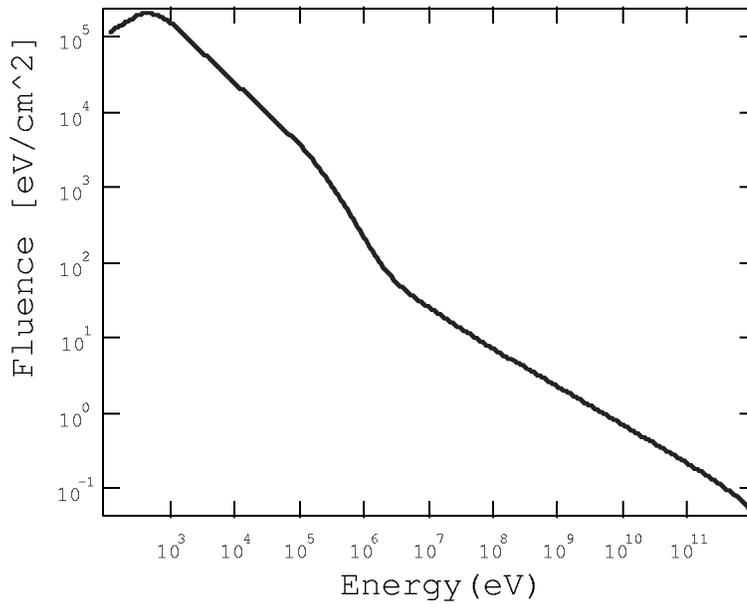


Figure 1: *Calculated flux for a simulated GRB located at $z = 1$ with a wind luminosity of 10^{53} ergs. The Lorentz factor are randomly chosen in the range $100 \div 1000$*

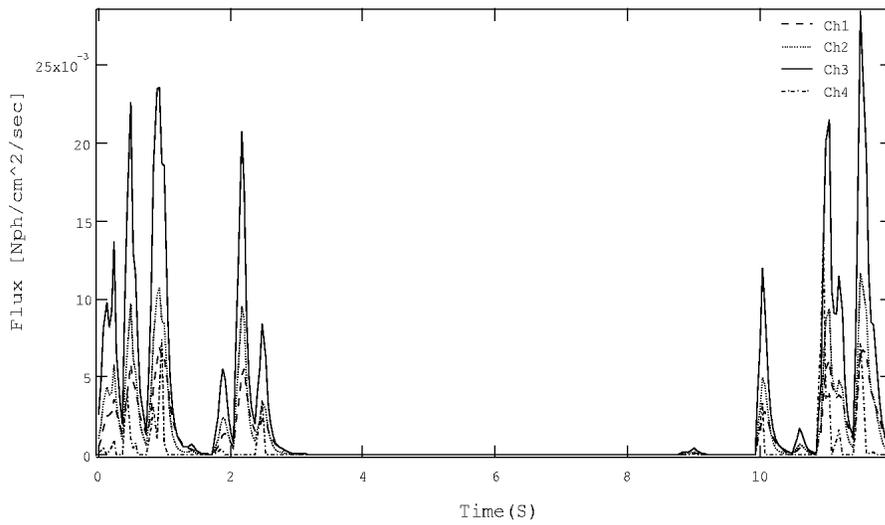


Figure 2: *Light Curves for four different channels. Ch1 corresponds to (20÷50KeV), Ch2 is (50÷300KeV), Ch3 is equal to (25KeV÷1.8MeV), while Ch4 is the LAT energy range:(20MeV 300GeV)*

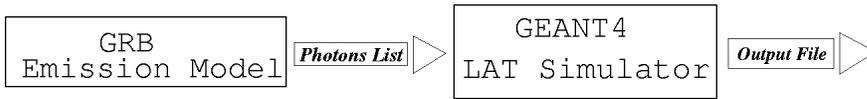


Figure 3: *Chart of the simulation chain. The emission model calculates the emission of a GRB source. A photons list is given as input to the simulation of the detector that compute the interaction of the photons in the detector. The photons list contains information on the spectral and temporal variability features of the emitted radiation. The output file could be used to reconstruct the initial signal and to study the capability of the detector.*

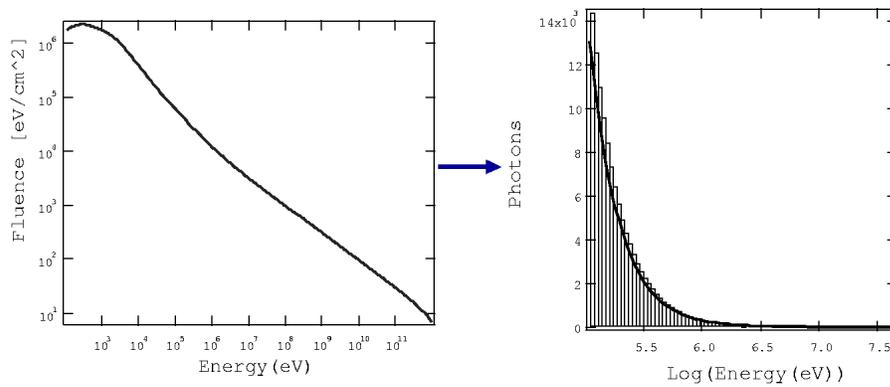


Figure 4: *The flux calculated from the simulation is used to extract photons at different energies. The right plot shows the histogram filled by the photons number for each energy bins for energies greater the 100KeV. The line is the photon number expected from the simulation.*

AUGER OBSERVATORY: THE WORLD'S LARGEST COSMIC RAY DETECTOR

C. Aramo for the Pierre Auger Collaboration
Sezione INFN and University of Catania

ABSTRACT

The Auger experiment is an international project devoted to the study of the cosmic rays particles with energies exceeding the predicted GZK spectral cutoff. The whole experiment will consist of two Observatories, in the Northern and Southern hemispheres. An engineering array, expected to be completed by the end of the year and consisting of 1/40 of the Southern Observatory, is under construction in Malargue, Argentina. According to the project, by the end of 2004, the Southern Observatory will consist of 1600 particle detectors covering an area of 3000 km². The Auger Observatory has been designed to work in a hybrid detection mode: a surface detector giant array (SD) will measure the lateral and temporal distribution of shower particles at ground level, and an air fluorescence detectors (FD) will measure the longitudinal development of the shower in the atmosphere above the SD. The concept of the experiment as well as the current status will be described.

1 Introduction

The puzzle set by the existence of cosmic rays with energies above 10^{20} eV, which may be an indication of new physics or exotic particles, is at present one of the hot topics in high energy astroparticle physics ^{1, 2, 3, 4}). The origin of these extremely high energy cosmic rays (EHECR) is very puzzling especially because of the well-known GZK (Greisen-Zatsepin-Kuzmin) ⁵) effect: if the cosmic rays are extragalactic, then a sharp cutoff at around several times 10^{19} eV in the observed spectrum is expected according to the energy degradation of the cosmic ray particles through the interaction with the microwave background radiation photons . This process limits the distance of the sources of particles with energies above 10^{20} eV to less than 100 Mpc from the Earth ^{6, 7, 8}). In the Fig. 1, where the updated AGASA measurement are displayed ²), one has a clearer view of what can be expected from a cosmological uniform distribution of conventional sources and what is observed. The GZK cutoff is clearly visible on the dashed line while the data suggests a change of slope as if a new phenomenon was rising above a steeply falling spectrum. The cutoff, that would be expected if the sources were cosmologically distributed and if the observed cosmic rays had no exotic propagation or interaction properties, is not present in the observed data. To test the different models that try to explain the absence of cutoff, it is crucial an experiment able to provide high quality and statistically significant data at the upper-end of the cosmic ray spectrum with good energy and angular resolution, high sensitivity to the composition and uniform exposure over the whole sky.

The Pierre Auger Observatory has been designed as a hybrid detector, taking advantage of complementary observational techniques - a surface array of water Cherenkov detectors coupled with four air fluorescence detector sites. Such a coupling provides excellent reconstruction of the shower axis geometry and estimation of the primary energy by using information from both the fluorescence and surface detectors, even at energies below the nominal detector threshold of 10^{19} eV. It also allows for powerful cross-checks of detector performance and analysis techniques through the comparison of surface only and fluorescence only reconstruction of the same showers. In this paper the concept of the experiment as well as the current status will be described.

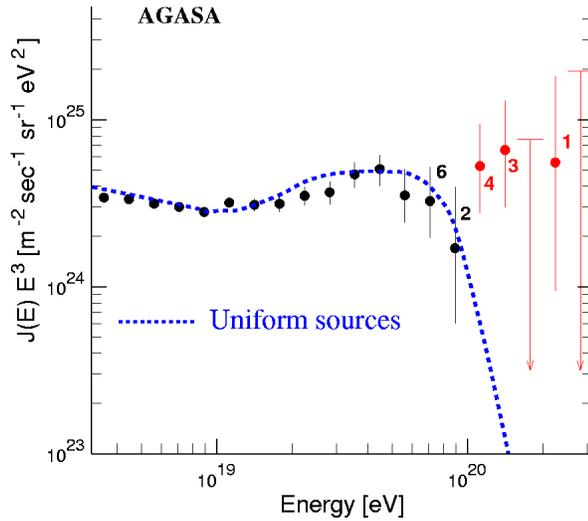


Figure 1: *Highest energy region of the cosmic ray spectrum as observed by the AGASA detector*²⁾. The numbers near the data points indicate the number of events in the corresponding energy bin. The arrows show 90 % confidence level upper limits. The dashed line is the expected spectrum if the sources were cosmologically distributed.

2 Surface Array and Fluorescence Detectors

Two similar observatories, covering an area of 3000 km² each, will be constructed in both hemispheres, in order to get a full sky coverage. The first one is being built in Malargue, Province of Mendoza, Argentina, and the second one could be located in Millard County, Utah, USA, at a later stage. The Pierre Auger apparatus is of hybrid type, as it combines two complementary techniques: a surface detector (SD) array will detect cosmic ray showers that hit the ground, and a fluorescence detector (FD) system will observe the nitrogen fluorescence light produced by the shower in the atmosphere above the SD array. Surface array stations are water Cherenkov detectors spaced 1.5 km forming an hexagonal grid (Fig. 2). Each detector is a cylindrical tank of 10 m² top surface and 1.2 m height, filled with filtered water and lined with a highly reflective material⁹⁾. The Cherenkov light is detected by three PMTs installed on the top. These stations will operate on battery-backed solar power and will communicate with a central station by using wireless radio links. Event timing will be provided through GPS (Global System Position) receivers. The Observatory is completed with fluorescence detectors: three "eyes" (6 telescopes each) will be installed at the periphery of the array and one at the center (12

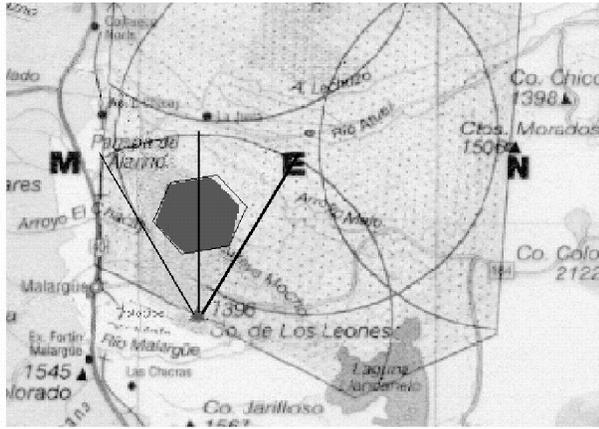


Figure 2: *Layout of the Observatory, where 1600 surface detectors will be deployed in the dotted area and the telescope systems in the marked circles at Los Leones, Coihueco, Morados at the site periphery and a fourth one at the center. The black hexagonal box is the area with 40 tanks of the EA. The lines represent the field of view (30°) of the first two fluorescence telescopes ¹⁰⁾.*

telescopes). A telescope consists of 440 hexagonal phototubes housed in a camera body designed to be fitted in the focal surface of a mirror of 3.4 m radius of curvature with field of view of about $30^\circ \times 30^\circ$ ¹⁰⁾. The main elements of the telescope ¹¹⁾ are shown in the Fig. 3 where, from left to right, we see the external shutter, the aperture system with the circular diaphragm and the UV transmitting filter, the camera and the large spherical mirror. The telescope is mounted inside a light tight enclosure. In the hybrid mode, the Pierre Auger Observatory is expected to have a 10 % energy resolution and an angular precision of 0.35° for energies above 10^{20} eV. For the surface array running alone those numbers became 12 % and 0.6° . Each observatory will have an aperture of $7400 \text{ km}^2 \text{ sr}$. The expected number of events at zenith angle less than 60° is ~ 5100 events above 10^{19} eV and ~ 60 events above 10^{20} eV per year.

3 The Engineering Array

Before proceeding to the construction of the full-size observatory, a subset of it, the "Engineering Array" (EA), is being built at the southern site near Malargüe in Argentina. It consists of an array of 40 surface detectors and two fluorescence telescopes. The goal of this small scale observatory is to test the individual components and their integrated operation, including the communications and data acquisition systems, in order to optimize the design, overcome

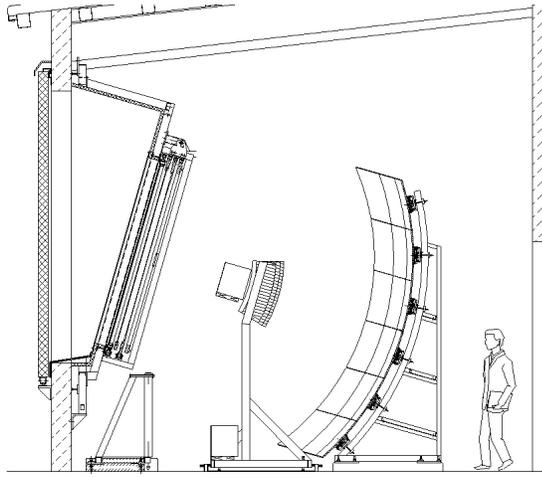


Figure 3: *The Auger Fluorescence Detector telescope.*

assembly and deployment difficulties and reduce construction time and costs in the production phase. Furthermore, the acquisition of some preliminary data, including a few events recorded in hybrid mode, is expected. The 40 deployed tanks have been equipped with liners and filled with purified water delivered by the Auger water plant. The solar power system has been installed on all the detectors and one of them is in data taking mode since 9th June transmitting monitoring data. At the end of July all 40 tanks have been deployed into the field together with their PMTs. At the beginning of August the first shower has been recorded. In the fluorescence detector building at the Los Leones site (Fig. 2), two out of the 6 bays are being equipped with fluorescence telescopes. They are oriented in such a way that each telescope observes the atmosphere above 20 SD stations. The construction of two fluorescence telescopes will allow the observation of showers crossing the field of view of both telescopes, thus permitting a test of the performance at the interface and a direct comparison of their detection capabilities. The construction of the first telescope has been accomplished. The mirrors have been fitted to their mountings, the camera has been installed, equipped with the corresponding 440 PMTs and their electronics and aligned with respect to the reference point. All the components of the aperture (supports, protective glass, filter, corrector lens and shutters) are in place. At the end of May 2001, the first FD telescope is being commissioned and is recording its first atmospheric shower event candidates (Fig. 4) at a rate of approximately 5 per hour with the corrector ring in operation. Progress

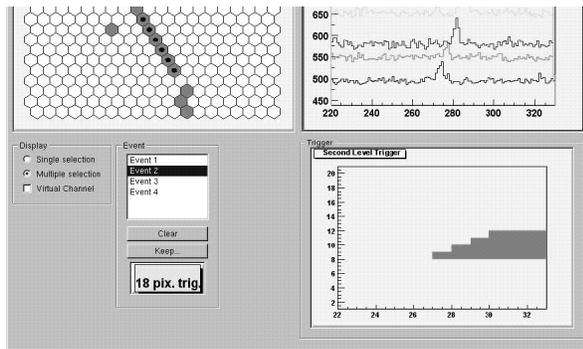


Figure 4: The on line shower event candidate as seen by the FD telescope. The picture shows on the upper left plot the pixel matrix with the triggered pixels in gray, on the upper right plot FADC traces for selected pixels and on the lower right the start of the trigger time for each row.

is underway with the installation of the calibration system and atmospheric attenuation monitors. The EA is expected to be completed and commissioned by the end of 2001 and the complete observatory by the end of 2004.

References

1. M.A. Lawrence *et al*, J. Phys. **G13**, 733 (1991).
2. N. Hayashida *et al*, Phys. Rev. Lett. **74**, 3491 (1994).
3. D.J. Bird *et al*, Astrophys. J. **441**, 144 (1995).
4. T.Abu-Zayyad *et al*, 26th ICRC, Salt Lake City, **Vol.3** (1999) p.264.
5. K. Greisen, Phys. Rev. Lett. **16** (1966); G.T. Zatsepin & V.A. Kuzmin, JETP letters **4** (1966).
6. F.A Aharonian and J.W Cronin, Phys. Rev. **D50**, 1892 (1994).
7. V.S Berezinsky, Yad.Fiz **11**, 339 (1970).
8. R.J. Protheroe and P.L. Biermann, Astropart. Phys. **7**, 181 (1997).
9. C.O. Escobar for the Pierre Auger Collaboration, 27th ICRC, Hamburg, **HE1.8** (2001), p.749.
10. Pierre Auger Design Report (1997)- www.auger.org.
11. G.Matthiae for the Pierre Auger Collaboration, 27th ICRC, Hamburg, **HE1.8** (2001), p.733.

(NEW) NEUTRINO PHYSICS AT GRAN SASSO NATIONAL LABORATORY

A. Bettini

Dipartimento di Fisica G. Galilei dell'Università di Padova; INFN Gran Sasso National
Laboratory and Sezione di Padova

ABSTRACT

Experiments in underground laboratories have shown evidence of physics beyond the standard model. The anomalies observed in electron-neutrinos from the sun and muon-neutrinos from cosmic rays interactions in the atmosphere can be explained if neutrino are massive and oscillate. The physics program at the INFN Gran Sasso Laboratory that we are defining will be focussed on the next phase of neutrino physics with a complementary set of experiments.

1. Introduction

Underground laboratories are complementary to those with accelerators in the basic research of the elementary constituents of matter and of their interactions and symmetries. They provide the low radioactive background environment necessary to the search for those extremely rare phenomena, which may give us information of the physics of extremely high energies. Indeed we have now for the first time strong hints for physics beyond the standard model. The evidence is in neutrino physics and has been obtained in underground laboratories, mainly Kamioka in Japan, Gran Sasso in Italy and, more recently, SNO in Canada.

On the basis of this evidence we know now that some of the assumptions of the Standard Model are not correct. Neutrinos have non-zero masses, electron neutrinos, muon neutrinos and tau-neutrinos - the particles produced by weak interactions and detected by our apparatuses - are not

the mass eigenstates and their flavour quantum numbers are not conserved. These findings point clearly to new physics. In the following, after a brief description of the Gran Sasso laboratory and a reminder of neutrino physics, I'll summarise the scientific program of the laboratory. Space does not allow here to review into details all the arguments; for more complete reviews the reader should consult^[1].

2. The Gran Sasso Laboratory (LNGS)

The INFN Gran Sasso Laboratories are located besides the freeway tunnel (10.4 km long) connecting L'Aquila and Teramo, at about 6 km from the west entrance, 120 km from Rome.

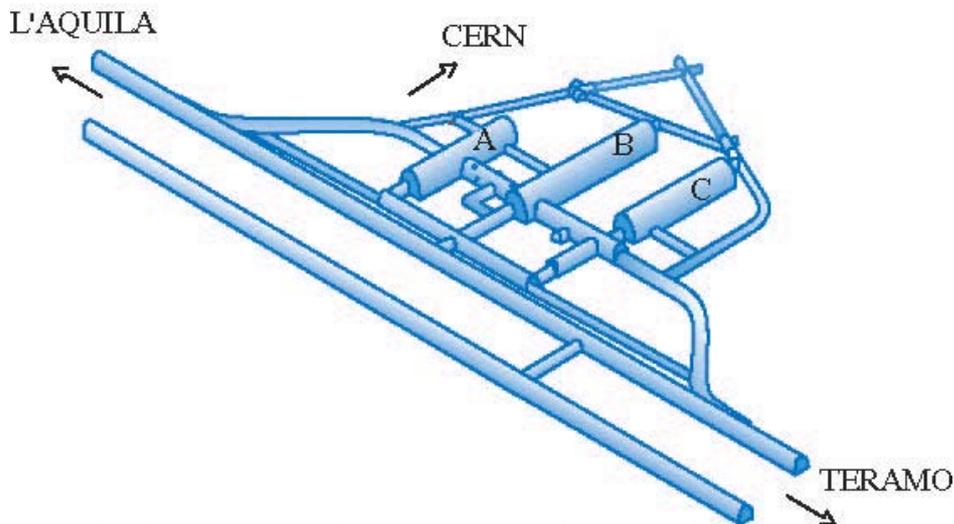


Fig. 1. Artist view of the underground facilities of the Gran Sasso Laboratory

Fig. 1 shows a view of the facility. The access is through the gallery of the freeway, allowing the transportation of large pieces of apparatus. The underground facilities consist of three experimental halls, called hall A, B and C, and a set of connecting tunnels and service areas, for a total surface of 18 000 m². The three halls are approximately 100 m long, 18 m wide, 18 m high. The infrastructures of the laboratory are completed by a number of buildings on the surface, near the western entrance of the tunnel, hosting offices, laboratories, shops, library, canteen, etc.

The flat shape of the massif, with an average rock overburden of 1400 m provides uniform coverage at all angles, giving a cosmic rays muon flux attenuation of a factor 10⁶. The neutron fluence from the dolomite rock is particularly low, 1000 times less than on the surface.

The mission of the laboratory is to host experiments in fundamental physics requesting very low levels of radioactive background and researches in other disciplines (notably geophysics and biology) that can profit of the unique environmental characteristics of the site.

A law approved by the Parliament in 1990 funds the completion of the Gran Sasso Laboratory with two new halls and with an independent access tunnel, necessary to guarantee a high safety standard. After a ten-year long delay, the new Government in 2001 has included the project in its public works programme as an “emergency” issue.

Experiments at Gran Sasso, in its little more than ten-year operational life, have already provided major discoveries and given important contributions to science^[2]. The first generation experiments, or at least some of them, are reaching or have reached completion. Taking office in 1997, I charged the international Scientific Committee of the Laboratory to examine in depth all the running experiments in order to determine on scientific grounds the data taking time still necessary to each of them to be completed. The experiments had been approved, in fact, without defining their overall occupation time of the laboratory underground space.

The review led to the conclusion that in the year 2001 almost half of the laboratory space would be available to new experiments. The knowledge of availability of space has stimulated the scientific community and a number of very interesting ideas and proposals have been submitted to the Laboratory. It is now clear that first class opportunities are present for the next experimental phase that may, with a bit of fortune, lead to major discoveries of physics beyond the present theory of elementary particles.

Neutrino physics will be the principal, but not the only issue of the research program for the next years. Experiments both with naturally produced neutrinos (from the Sun, from the atmosphere and from Supernova explosion) and artificially produced ones (mainly from CERN, but possibly by other sources too) are being built or planned. Other experiments will try to understand the nature of the electron neutrino and search for the Majorana mass; still others (not reviewed here) will continue with increased sensitivity the search for non-baryonic dark matter. The measurements of thermonuclear cross-sections at energies relevant for the stars and Sun combustion processes will continue with an improved underground accelerator facility.

3. Neutrino masses and mixing

Experiments in underground laboratories have provided strong evidence for neutrino oscillations. Two and independent are the physical sources of this information: the electron neutrinos from the Sun and the muon neutrinos indirectly produced by cosmic rays in the atmosphere. The corresponding two oscillation phenomena take place with very different periods, inversely proportional to the differences between the relevant

mass eigenstates. I'll call m^2 and m'^2 the square mass differences for the solar and atmospheric oscillation respectively.

The neutrino states with definite flavour (ν_e , ν_μ and ν_τ), those produced by weak interactions and detected by our instruments, are linear combinations of the mass eigenstates (ν_1 , ν_2 and ν_3),

$$\nu_l = \sum_{i=1}^3 U_{li} \nu_i$$

where $l = e, \mu, \tau$. The mixing matrix being unitary, its elements can be expressed in terms of four independent real parameters. These are usually taken as three "mixing angles" (θ_{12} , θ_{13} and θ_{23}) and a phase factor. The phase factor gives CP violating effects in the lepton sector, extremely important, but unfortunately still very far to be experimentally accessible. As a consequence, I will, for simplicity, forget it and consider only real matrix elements. Two further phases, α and β , irrelevant for oscillations, are present if neutrinos are Majorana particles. We have

$$\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} \\ 0 & 1 & 0 \\ -s_{13} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & -s_{12} & 0 \\ s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The situation is now much more complex than in the case of only two neutrino states. Two different oscillations take place with different frequencies or, in other words, at different flight times. The expressions for the probability to observe a state of definite flavour are much more complicated than in the two-flavour case. Just to give an example, an approximate expression of the probability to observe a ν_e in an initially (monochromatic) ν_μ beam propagating in a vacuum is

$$P_{\nu_\mu \rightarrow \nu_e} = \sin^2(\theta_{23}) \sin^2(2\theta_{13}) \sin^2\left(1.27 \frac{m^2(\text{eV}^2) L(\text{km})}{E(\text{GeV})}\right)$$

This is not the complete formula, it is a good approximation for flight times relevant for the first oscillation when the second, slower one has not yet started. Notice that the probability amplitude depends now from two mixing angles. Notice also that the probability is different if one of the angles, θ_{23} , is in the first or second octant. Considering all the cases one sees that the full range $0 < \theta_{12}, \theta_{13}, \theta_{23} < \pi/2$ must be considered, and not $0 < \theta < \pi/4$ as is still frequently, but wrongly, done. The variable $\sin^2 2\theta$ is misleading and should not be used. Better variables are $\sin^2\theta$ or $\tan^2\theta$ or just θ . In practice I'll be sometimes forced to use $\sin^2 2\theta$ in the following, when quoting results presented in this form, but only in cases safe from errors.

A summary of the present knowledge is the following.

Electron neutrinos are produced by the thermonuclear processes in the core of the Sun. When the ν_e flux is measured on the Earth, substantially lower values than expected are found. A fundamental contribution was given by the GALLEX experiment at Gran Sasso. It measured for the first time the ν_e flux from the pp reaction, that is model independent and known from solar luminosity (2% uncertainty). GALLEX was also the first radiochemical experiment to be absolutely calibrated with an artificial ν_e source. Other important information is provided by the SuperKAMIOKANDE measurement of the high-energy (>5 MeV) neutrino spectrum both during the day, when neutrinos reach the detector directly, and during the night when they cross the Earth (possible matter effects). The experimental evidence can be explained only if neutrinos behave in a non-standard way, the simplest hypothesis being oscillations (including MSW effect). This phenomenon depends mainly on two parameters, the square mass difference m^2 and the mixing angle θ_{12} . Solar neutrino data do not select a unique solution, but are compatible with a few, amongst which future experiments will choose. Notice that all but one solutions (SMA, that is disfavoured by data) are close or equal to "maximum mixing" meaning here that $|U_{e1}|^2 = |U_{e2}|^2 = 1/2$.

The second anomaly has been convincingly observed by Super-Kamiokande and confirmed by MACRO at Gran Sasso in the "atmospheric" neutrinos. The simplest interpretation is we are observing a second oscillation phenomenon mainly between ν_μ and ν_τ . The square mass difference, as measured by Super-Kamiokande, is in the range $1.5 \times 10^{-3} \text{ eV}^2 < m^2 < 5 \times 10^{-3} \text{ eV}^2$. The mixing is compatible to be maximum, meaning now that $\theta_{23} = 45^\circ$ or equivalently that $|U_{\mu 2}|^2 = |U_{\mu 3}|^2 = 1/2$.

Finally the electron antineutrinos disappearance CHOOZ^[3] experiment gives the limit $|U_{e3}|^2 < 0.03$. In conclusion, two of the mixing angles, θ_{12} and θ_{23} appear to be close to 45° (maximum mixing), while the third, θ_{13} is close to zero. The pattern is very different from that of quarks.

From these pieces of evidence we can assume that the neutrino mass spectrum consists of two nearby levels, m_1 and m_2 , and a third more separated one, m_3 . The smaller mass difference $\delta m^2 = m_2^2 - m_1^2$ is responsible of the solar anomaly, the larger one $\Delta m^2 = m_3^2 - m_2^2 = m_3^2 - m_1^2$ of the atmospheric one. In other words the neutrino mass spectrum is composed of a doublet of states very close together and of a third, more separate state. The last one is a superposition of ν_μ and ν_τ almost one to one, with possibly a small ν_e component.

As neutrino oscillations depend on the absolute value of the difference between the squares of the masses, we do not know whether the third state is higher ($m^2 > 0$, called "normal" spectrum) or lower ($m^2 < 0$, called "inverted") than the doublet. Neither we know the absolute scale of the masses. The spectrum may be degenerate, when the three masses are

almost equal, or hierarchical, when the masses of the doublet are of the order of the square roots of the two square mass differences.

Notice that mass is a property of the stationary states (the eigenstates) and that talking of ν_e , ν_μ or ν_τ mass is improper and in some cases misleading. What is meant depends in fact on what and how one measures (or limits).

Consider as an important example the limits on the “electron-neutrino mass” $\langle m_{\nu_e} \rangle$ that are obtained by measuring the electron energy spectrum in the Tritium beta decay. If neutrinos are massive, the spectrum should show three steps in correspondence with the three masses. But these cannot be resolved and one measures an average effect

$$\langle m_{\nu_e}^2 \rangle = |U_{e1}|^2 m_1^2 + |U_{e2}|^2 m_2^2 + |U_{e3}|^2 m_3^2$$

Presently two experiments give the upper limit $\langle m_{\nu_e} \rangle < 2.3 \text{ eV}^{[4]}$.

If neutrinos are massive Majorana particles, a very rare process, the neutrino-less double beta decay ($0\nu 2\beta$) can happen in some nuclides. No positive signal has been observed and limits on the corresponding lifetimes have been set. From each of them a limit on the electron-neutrino “effective mass” M_{ee}^M can be extracted, taking into account the relevant nuclear matrix elements. The corresponding uncertainties are typically a factor two. As a consequence, it is mandatory for a complete research program to include different double-beta active isotopes in the search. In the present case the “mass” that is measured, or limited is the quantity

$$|M_{ee}^M| = \left| |U_{e1}|^2 m_1 + |U_{e2}|^2 e^{2i\alpha} m_2 + |U_{e3}|^2 e^{2i\beta} m_3 \right|$$

Notice that cancellations can happen due to the phase factors. Presently the best limit, $M_{ee}^M < 340 \text{ meV}$ (90% c.l.) is given by the Heidelberg-Moscow^[5] experiment at Gran Sasso, obtained with a 37.2 kg x yr exposure of an enriched ^{76}Ge detector.

The same group has proposed in 1997 the GENIUS^[6] experiment aiming for a forward jump in the sensitivity with a large increase in the enriched Ge mass (1000 kg) and a drastic reduction of the background. Naked enriched Ge crystals would be immersed in a liquid N_2 bath, 10 m across used both for cooling the crystals and to screen the external radioactivity. The experience of BOREXINO shows that extremely low radiopurity levels can be reached liquid N_2 and Monte Carlo calculations show that the technique should allow to reduce the background, in the relevant energy, to reach $b = 3 \times 10^{-4} \text{ events/(kg keV yr)}$. This would allow the experiment to reach the 10 meV neutrino mass range. To prove that such a large reduction in the background is possible in practice Monte Carlo calculations are not enough and a series of tests is necessary. To this aim the GENIUS-TF^[7] proposal, based on 40 kg of natural Germanium, has been approved.

The most sensitive experiment on a different isotope is MIBETA, again at Gran Sasso, with 20 TeO_2 crystals operated as bolometers at cryogenic

temperatures. The total detector mass is almost 7 kg of natural Te or of 2.3 kg of the double-beta active ^{130}Te isotope. MIBETA has reached an exposure of 3.3 kg yr with a background level $b = 0.6 \text{ ev}/(\text{kg keV yr})$, giving the limit $M_{ee}^M < 2 \text{ eV}$ ^[8].

The next experiment with the same technique is CUORICINO^[9] consisting of 56 TeO_2 crystals, 0.76 kg each, corresponding to a total ^{130}Te mass of 14.3 kg. The first crystals are in the test phase. If the background level will be reduced at $b = 0.1 \text{ ev}/(\text{kg keV yr})$, as it appears to be feasible from the results of the tests, sensitivity around 400 meV will be reached in M_{ee}^M . Further increase in the mass, by an order of magnitude, and drastic reduction of the background are being studied in view of the CUORE project aiming to a 50 meV sensitivity. It will consist in 1000 natural Te crystals equal to those of CUORICINO with a sensitive ^{130}Te mass of 250 kg and aim to a background rate $b = 10^{-2} \text{ events}/(\text{kg keV yr})$.

In conclusion, double beta decay experiments could reach sensitivities in the range, $M_{ee}^M = 30 - 50 \text{ meV}$. These are extremely interesting values, being close to the square root of the atmospheric square mass difference, i. e. 40-70 meV. In case of a degenerate or of an inverted spectrum the sensitivity of GENIUS and CUORE might be enough to detect the signal.

4. The next steps

In the previous paragraphs I have briefly described the recent experimental findings that have given origin to the new neutrino physics. Clearly we might just be entered in a new field that can reserve complete new discoveries for the future. The program of the next years should include experiments able to

- observe oscillation signals both for the atmospheric anomaly and the solar one. In neither case we have yet observed a non ambiguous sign of oscillation. In both cases oscillations give the simplest explanation, but more exotic interpretations are not excluded.
- confirm the atmospheric neutrino oscillations with experiments with a neutrino beam produced at a far away accelerator. Composition (mainly μ) and energy spectrum of an artificially produced beam are under control. Both μ disappearance and appearance experiments are planned. The K2K experiment is running since 1999: a muon-neutrino beam is produced at the Tsukuba KEK Center and sent to the Super-Kamiokande detector 250 km away. The low neutrino energies (2-3 GeV) give good sensitivity even with low statistics^[10]. The NUMI program at Fermilab is building a muon-neutrino beam to shoot on the MINOS detector being built in the Soudan mine 730 km away in Minnesota. The experiment is now planned to start data taking in 2005 in a disappearance mode^[11].
- discover if the flavour into which the atmospheric ν_μ oscillate is indeed ν_τ or else with a ν_τ appearance experiment, as planned by the CNGS project

in Europe. This issue is clearly connected with the existence of low-mass sterile neutrinos coupled to known particles.

- improve the knowledge of the mixing parameters.
- measure the sign of Δm^2 . Is the spectrum “normal” or “inverted”? We have already recalled the chances for the $0\nu 2\beta$ search and Supernova neutrinos at this purpose. Others exist for a sign sensitive experiment on atmospheric muon neutrinos.
- improve the knowledge of δm^2 . Choose solar solution.
- detect, if any, the $\nu_\mu \rightarrow \nu_e$ oscillation at Δm^2 . Is $U_{e3} = 0$?
- search for CP violation in the lepton sector. This is extremely difficult, for a number of reasons, including the fact that the effects are suppressed by the factor $m^2 / m^2 \ll 1$ and by the smallness of $|U_{e3}|$
- determine the nature of neutrinos: Majorana or Dirac?
- measure the absolute values of the masses

Experiments at Gran Sasso Laboratory can give important contributions, provided that we will be able to build a coherent program. This will be done on the basis of the many interesting ideas and proposals that have been submitted and that are in different stages of development and of the resources that will become available. In the following I'll briefly describe these proposals.

5. Neutrinos from CERN

An important component of the program will be the CNGS project. An artificial, well-controlled neutrino source will be built at CERN for experiments at LNGS. Both the beam^[12] and the experiments will be optimised for ν_τ appearance, where ν_τ 's are observed through the process

$$\nu_\tau + N \rightarrow \tau^- + N'$$

The ν_τ appearance probability in an initially pure ν_μ beam of energy E is

$$P_{\nu_\mu \rightarrow \nu_\tau} = \sin^2(2\theta_{23}) \cos^4(\theta_{13}) \sin^2 \left(1.27 \frac{m^2(\text{eV}^2) L(\text{km})}{E(\text{GeV})} \right)$$

As $\theta_{13} = 0$, we have approximately

$$P_{\nu_\mu \rightarrow \nu_\tau} = \sin^2(2\theta_{23}) \sin^2 \left(1.27 \frac{m^2(\text{eV}^2) L(\text{km})}{E(\text{GeV})} \right)$$

In this case the commonly used two-neutrino formalism is justified, only as long as we forget the minority muon neutrino to electron-neutrino oscillation.

Running in the “shared” mode, the beam will give 3200 CC ν_μ interactions per year in a kiloton fiducial mass detector at LNGS

corresponding to 25 ν_τ interactions for $m^2 = 3.5 \times 10^{-3} \text{ eV}^2$ and maximum mixing (1.7 times more in dedicated mode operation). The charged daughters of τ 's will be detected, in one or more decay channels: $\tau^- \rightarrow \mu^- \nu_\mu \nu_\tau$ (18%); $e^- \nu_e \nu_\tau$ (18%); $h^- \nu_\tau n\pi^0$ (50%); $2\pi^- \pi^+ \nu_\tau n\pi^0$ (14%). Two main background rejection tools are available: 1. the direct observation of τ decays requiring micrometer scale granularity and sub-micron resolution, which are possible only by the emulsion technique (OPERA); 2. the use of kinematic selection, which requires good particle identification and good resolution in momentum unbalance (ICARUS). ICARUS^[13] is a liquid argon time projection chamber providing bubble chamber quality 3D images of the events, continuous sensitivity, self-triggering capability, high granularity calorimetry and dE/dx measurement. The R&D program performed between 1991 and '95 on a 3 t detector solved the major technical problems with the detector continuously running for several years, showing the reliability of the technique. The technique was then developed for the industrial production of a kiloton size detector. Its structure will be modular. A module has a mass of 600 t (T600) and is composed of two 300 t units, transportable on highways. The units will be completely assembled and tested before being separately transported to Gran Sasso. A 300 t unit has been completed and successfully operated in summer 2001. Fig. 2 shows the superior quality and the richness of information provided by the technique. The safety issues connected with the installation of a large cryogenic volume underground are also being studied. The project foresees the construction of a series of T600's to cover a broad physics program, including ν_τ appearance in the CNGS program.

Fig. 2 A cosmic ray event in the first 300t ICARUS module

The design of OPERA^[14] combines in its basic cell, shown in Fig. 3, the high precision tracking capability of nuclear emulsion and the large target mass given by lead plates (1mm thick). The basic building block of the target structure is a “brick”, a sandwich of contiguous cells enclosed in a light-tight evacuated envelope. A wall is followed by electronic trackers with moderate resolution with the scope to identify the brick where a neutrino interaction took place and to guide the off-line scanning. Fired bricks will be removed and processed (alignment, development and scanning of the emulsion sheets) on a day by day basis.

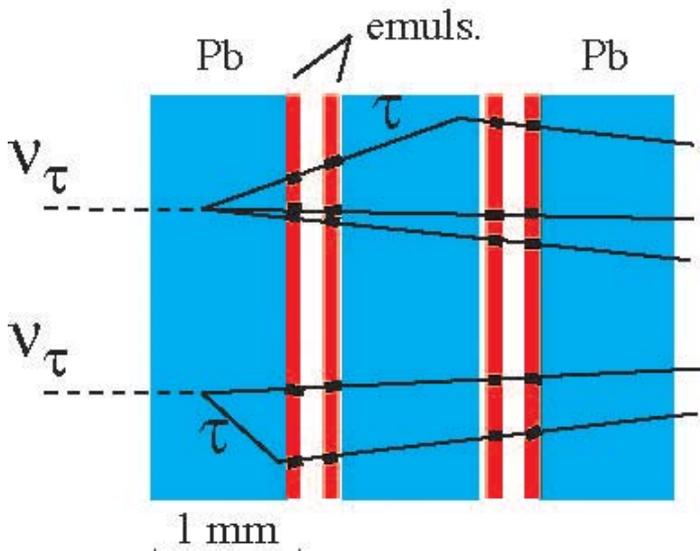


Fig. 3. The OPERA basic cell structure. Two types of events are shown and the principles for detection

6. Atmospheric neutrinos

Atmospheric neutrino experiments are complementary to CNGS. The two principal aims of the MONOLITH proposal^[15] are the observation of the oscillation pattern and the accurate measurement of m^2 . Notice that both aims are easier if m^2 is lower.

The oscillation probability is a periodic function of the L/E variable, the ratio between the muon-neutrino energy E and its flight length L . L is obtained from the neutrino direction, inferred from that of the μ . To have a good correlation, one must use only μ 's above a GeV or so, were unfortunately the cosmic rays flux is low. As a consequence several kiloton mass detectors are needed but with coarse resolution. MONOLITH is a 35 kt spectrometer made of 8 cm thick horizontal Fe magnetised

plates. The interleaved tracking planes have 1 cm spatial resolution and good (1 ns) timing, for up/down discrimination.

For a given direction, down going ν_μ 's do not oscillate, while upward going do. The ratio between the two fluxes is known with small systematic uncertainty. The distribution of this ratio as a function of the zenith angle (i.e. L/E) will show the oscillation pattern (with 100 – 150 kt yr exposures) and precisely determine m^2 .

7. Solar neutrinos

GALLEX has been concluded, but its improved version, GNO, is running and has published results corresponding to 35 solar runs (each one-month long). The experiment aims to reduce the systematic (now = 4.6%) and statistical uncertainties well below 5%. Increase of the sensitive mass by a factor two to three is foreseen in the proposal. If this will not be feasible, the date of closure of the experiment (a few years from now) will be defined on scientific grounds.

The measurement of the mono-energetic, 0.86 MeV, Be neutrino flux in real time is the principal aim of the BOREXINO^[16]. Indeed Be neutrinos flux appears to be particularly sensitive to oscillation parameters. Electrons resulting from a neutrino (any flavour, but ν_μ and ν_τ with smaller than ν_e cross-sections) scattering in the liquid scintillator detector medium will produce a light flash that will be detected by photomultipliers. 300 t of ultra-pure pseudocumene will be contained in a nylon sphere, the 100 t innermost mass being the sensitive volume. A larger volume of pseudocumene inside a 13.7 m diameter stainless steel sphere hosting the optical modules surrounds the nylon sphere. This sphere is immersed in a 2500 t purified water tank.

The experiment is designed with a threshold of 0.25 MeV. The main problem at such low energies is the control of the background due to the always present radioactive isotopes. An intense R&D program has been carried out in the last ten years to select materials and to purify them at unprecedented limits of radio-purity. In parallel, techniques have been developed to measure ultra-low levels of radioactivity. Record levels of $10^{-16} - 10^{-17}$ (g of contaminant/g of material) for ^{232}Th and ^{238}U have been achieved.

To complete program on solar neutrino physics we need to measure in real time the neutrino spectrum in order to separate the contributions of the different branches, pp , ^7Be and ^8B with a flavour sensitive experiment. The LENS^[17] proposal addresses the problem using ν_e capture by ^{176}Yb nuclides that go into an excited ^{176}Lu state. The Yb is loaded into an organic liquid scintillator to detect the electron resulting from the capture and the delayed γ , used as a tag, from the excited Lu decay. Notice that the ^{176}Lu ground state is higher than that of ^{176}Yb , making this nuclide stable

against beta decay. This characteristics and the low (301 keV) threshold for neutrino capture make ^{176}Yb practically unique.

The techniques needed to prepare large quantities of scintillator doped with a large fraction of Yb (at least 8%), with a reasonable light yield, with a good attenuation length (several metres) and chemically have been developed. The R&D are continuing to obtain requested radiopurity levels. Neutrino sources necessary for calibration must be procured.

8. Nuclear astrophysics

The solar model calculations need the values of the cross sections of the nuclear reactions involved in the different branches of the pp cycle. These cross sections are so small, due to the extremely low Coulomb barrier penetration probability at the relevant energies (called the Gamow peak), that their measurement became only recently possible in the low background Gran Sasso environment.

The LUNA experiment, based on a 50 kV ion accelerator, has already measured the cross section of the important $^3\text{He} + ^3\text{He} \rightarrow 2p + ^4\text{He}$ reaction down to 17 keV^[18] (where the cross section is only 20 fb and the rate 2 events/month!) below the Gamow peak. No resonance is present.

LUNA2 is the second-generation experiment. A 400 kV accelerator has been designed and installed. It is now in operation, with beam energy resolution better than 70 eV and long-term stability of 10 eV. A BGO-4 - summing detector completes the new facility. The gas target is located inside a borehole of the detector. A good energy resolution is indeed essential to reduce the background. The reactions $^{14}\text{N} (p,\gamma) ^{15}\text{O}$, $^3\text{He} (^3\text{He},\gamma) ^7\text{B}$ and $^7\text{Be} (p,\gamma) ^8\text{B}$ will be studied.

9. Neutrinos from Supernovae

Type II Supernovae in the Galaxy or in the Magellanian Cloud produce enough neutrinos (all flavours) to be observable. Notice that when leaving the Supernova core, electron neutrinos and antineutrinos have a softer spectrum (average energy approximately 10 MeV, I'll call it "soft") than the other flavours (average energy approximately 20 MeV, I'll call it "hard"). Neutrinos then cross the mantle, a medium of very high density in which important matter induced flavour conversions take places. Having left the star, the mass eigenstates propagate, independently one from the other, to our detector, possibly with slightly different velocities. Clearly the flux of a flavour we measure may be extremely different from that produced in the Supernova core. The measurement of the arrival times cannot, neither in principle, measure or limit, as is frequently but wrongly claimed, the mass of the detected neutrino flavour, for example, of the tau neutrino.

On the other hand we can extract information on neutrino mixing. To make an example, consider the case in which $|U_{e3}|^2$ is not too small ($>$ a few 10^{-4}). It can then be seen^[19] that if $m^2 > 0$ the electron neutrino spectrum as detected on Earth is equal to the originally produced muon and tau neutrino spectra (that are equal), it is then hard. On the other hand, the electron antineutrino spectrum is halfway between the soft and the hard ones. If $m^2 < 0$, the roles of electron neutrinos and antineutrinos are inverted. The measurement of the ratio of neutrino and antineutrino spectra, which is almost model-independent, would allow determining the sign of m^2 .

At Gran Sasso the LVD dedicated experiment has a 1080 t sensitive mass of organic liquid scintillator. The detector has a modular structure consisting of 912 tanks each seen by three photomultipliers. The tanks are read out independently, so allowing a very high up time, at least of a part of the apparatus (99.3% in 2000). LVD is mainly sensitive to electron-antineutrinos through the process $\bar{\nu}_e + p \rightarrow n + e^+$ (a few hundreds events for a collapse in the Galaxy centre), but also to electron neutrinos through $\nu_e + {}^{12}\text{C} \rightarrow e^- + {}^{12}\text{N}$ and to antineutrinos through $\bar{\nu}_e + {}^{12}\text{C} \rightarrow e^+ + {}^{12}\text{B}$. These last processes have thresholds around 15 MeV. This fact excludes a large fraction of the soft spectrum, not of the hard. As a consequence the expected yield increases by an order of magnitude if electron neutrinos or antineutrinos are made harder by matter effects.

12. Conclusions

Neutrino physics has entered in the last years a new age. The discovery of neutrino oscillations has shown that neutrinos have non-zero masses and that the leptonic flavour numbers are not conserved. The search for neutrino-less double beta decays has already reached sensitivity in Majorana mass capable to limit or to contradict some high energy extensions of the standard theory. Dark matter searches are reaching the regions where signals might appear.

These results that point to physics beyond the standard theory have been obtained in underground low background laboratories. Gran Sasso has contributed, as I have discussed.

An extremely interesting future appears to be in front of us where revolutionary discoveries might become possible. In particular the space that will be soon available at Gran Sasso and the quality of its infrastructures have stimulated many interesting ideas and proposals. These are in different stages of research and development, of test and of preparation. Presumably not all of them will become a running experiment, but we have good chances that at least a few will, with a bit of fortune, produce in the next years outstanding results.

References

- [1] A. Bettini. Physics Beyond the Standard Model at the Gran Sasso Laboratory, *USPEKHI* **171** (2001) 977-1003; A. Bettini, *New Neutrino Physics*, *Rivista del Nuovo Cimento*, Nov. 2001
- [2] A. Bettini. The Gran Sasso Laboratory 1979-1999. INFN 1999
- [3] CHOOZ Collaboration. M. Apollonio et al. *Phys. Lett.* **B466**, 415 (1999)
- [4] C. Weinheimer et al. *Phys. Lett.* **B460** (1999) 219; NEUTRINO 2000, *Nucl. Phys. B (Proc. Suppl.)* **91** 273 (2001); V. M. Lobashev et al. *Phys. Lett.* **B460** (1999) 219; NEUTRINO 2000, *Nucl. Phys. B (Proc. Suppl.)* **91** 280 (2001)
- [5] H.V. Klapdor-Kleingrothaus, "Sixty Years of Double Beta Decay – From Nuclear Physics to Beyond Standard Model Particle Physics", World Scientific, 2000
- [6] LNGS-LOI 9/97; LNGS-LOI 9/97 add. 1; LNGS P23/2000; MPI-Report MPI-H-V26-1999
- [7] MPI-Report MPI-H-V4-2001; LNGS P27/2001
- [8] A. Alessandrello et al: *Physics Letters B*, **486**, (2000) pp. 13-21.
- [9] LNGS-LOI 16/99; LNGS-LOI 12/98
- [10] K2K Collaboration at NEUTRINO 2000. *Nucl. Phys.* **B 91** (2001) 203
- [11] MINOS Technical Design Report, Nu-MI-L-337, October 1998
- [12] CERN 98-02, INFN/AE-98/05 and CERN-SL/99-034(DI), INFN/AE-99/05
- [13] ICANOE proposal to LNGS-SC and CERN SPSC: LNGS-P21/99, CERN/SPSC 99-25 and LNGS-P21/99.Add.1 and 2, CERN/SPSC 99-40
- [14] CERN/SPSC 2000-028; SPSC/P318; LNGS P25/2000. July 10, 2000
- [15] LNGS P26/2000; CERN/SPSC 2000-031; SPSC/M657. August 15th, 2000
- [16] C. Arpesella et al. BOREXINO proposal (Univ. of Milan) 1991; F. v. Feilitzsch et al. *Astro Phys.* **8** (1998) 141; BOREXINO Collaboration, hep-ex/0012030.
- [17] LNGS-LOI 18/99; LNGS P18/99 add.1
- [18] R. Bonetti et al.; *Phys. Rev. Lett.* **82**, 26, (1999), 5205-5208
- [19] G. Dutta, D. Indumathi, M. V. N. Murthy and G. Rajasekaran, *Phys. Rev. D* **61** (2000) 013009; T. K. Kuo and J. Pantaleone, *Phys. Rev. D* **37**, (1988) 298

PROBING GALAXY FORMATION WITH HIGH ENERGY GAMMA-RAYS

Joel R. Primack

Physics Department, University of California, Santa Cruz, CA 95064 USA

ABSTRACT

I discuss how measurements of the absorption of γ -rays from GeV to TeV energies via pair production on the extragalactic background light (EBL) can probe important issues in galaxy formation. My group uses semi-analytic models (SAMs) of galaxy formation, set within the CDM hierarchical structure formation scenario, to obtain predictions of the EBL from 0.1 to 1000 μ m. SAMs incorporate simplified physical treatments of the key processes of galaxy formation — including gravitational collapse and merging of dark matter halos, gas cooling and dissipation, star formation, supernova feedback and metal production — and have been shown to reproduce key observations at low and high redshift. We have improved our modelling of the spectral energy distributions in the mid-to-far-IR arising from emission by dust grains. Assuming a flat Λ CDM cosmology with $\Omega_m = 0.3$ and Hubble parameter $h = 0.65$, we investigate the consequences of variations in input assumptions such as the stellar initial mass function (IMF) and the efficiency of converting cold gas into stars. We also discuss recent attempts to determine the emitted spectrum of high

energy gamma rays from blazars such as Mrk 501 using the synchrotron self-Compton model and the observed X-rays, and note that our favorite SAM EBL plus the observed spectrum of Mrk 501 do *not* imply unphysical upturns in the high energy emitted spectrum — thus undermining recent claims of a crisis with drastic possible consequences such as breaking of Lorentz invariance. We conclude that observational studies of the absorption of γ -rays with energies from ~ 10 GeV to ~ 10 TeV will help to determine the EBL, and also help to explain its origin by constraining some of the most uncertain features of galaxy formation theory, including the IMF, the history of star formation, and the reprocessing of light by dust.¹

1 Introduction

The extragalactic background light (EBL) represents all the light that has been emitted by galaxies over the entire history of the universe. The EBL that we observe today is an admixture of light from different epochs, its spectral energy distribution (SED) distorted by the redshifting of photons as they travel to us from sources at different distances. It is therefore a constraint on both the intrinsic SEDs of the sources and their distribution in redshift. At present, there is more than a factor of two uncertainty in the amplitude of the EBL in the UV, optical, and near-infrared ²⁾. The EBL in the mid-IR is even more uncertain. The far-IR background measured at $\gtrsim 100\mu\text{m}$ ^{3, 4, 5, 6)} represents at least half of the total energy in the EBL, yet the sources that produced it remain uncertain.

High energy γ -ray astronomy promises to help resolve these uncertainties by providing independent constraints on the EBL, in the mid-IR with E_γ in the ~ 10 TeV energy range, and in the $0.1\text{-}3\ \mu\text{m}$ range with $E_\gamma \sim 100$ GeV via the new low-threshold instruments that will soon be available. High energy γ -rays from sources at cosmological distances are absorbed via electron-positron pair production on the diffuse background of photons that comprises the EBL. Thus, γ -ray observations of objects with known redshift and intrinsic spectral shape will constrain the EBL in these crucial wavelength regimes by measuring the optical depth of the Universe to photons of various energies. This in turn will help to constrain some of the most fundamental uncertainties in physical

¹This paper is an updated version of ¹⁾.

models of galaxy formation.

In order to illustrate this, in this paper we use a “forward evolution” approach, which attempts to model the essential features of galaxy formation using simple recipes. These semi-analytic models are set within the modern Cold Dark Matter (CDM) paradigm of hierarchical structure formation, and trace the gravitational collapse and merging of dark matter halos, the cooling and shock heating of gas, star formation, supernovae feedback, metal production, the evolution of stellar populations and the absorption and re-emission of starlight by dust. This machinery has been used extensively to predict optical properties of low-redshift galaxies, with good results (e.g., ^{7, 8}); reviewed and extended in ^{9, 10}, hereafter SP and SPF). A semi-analytic approach was also used by Devriendt and Guiderdoni ¹¹) to make predictions of counts and backgrounds in the mid-to-far-IR, with more detailed modelling of dust extinction and emission, but less detailed modelling of merging and star formation. We have now combined the strengths of these two approaches, by integrating the stellar SEDs and dust modelling of ^{12, 11}) into the galaxy formation SAM code of the Santa Cruz group.

Some parts of the “standard paradigm” of galaxy formation represented by our SAMs are relatively solid. For example, once a cosmological model and power spectrum are specified, it is straightforward to compute the gravitational collapse of dark matter into bound halos using N -body techniques, and analytic formalisms such as those used in our modelling ¹³) have been checked against these results ¹⁴). Within the range of values for the cosmological parameters allowed by existing observational constraints (i.e., $\Omega_{\text{matter}} \simeq 0.3-0.5$, $\Omega_{\text{matter}} + \Omega_{\Lambda} \simeq 1$, $H_0 \simeq 60-80$ km/s/Mpc; see e.g. ¹⁵) for a summary), these results do not change significantly. Similarly, modelling of gas cooling appears to be fairly robust and agrees well with hydrodynamic simulations ¹⁶). However, other aspects, notably the efficiency of conversion of cold gas into stars, the effect of subsequent feedback due to supernovae winds or ionizing photons, the stellar initial mass function (IMF), and the effects of dust, remain highly uncertain, and some predictions are quite sensitive to their details.

For example, SPF showed that the star formation history of the Universe and the number density of high redshift $z \gtrsim 2$ “Lyman-break” galaxies (LBGs; e.g. ¹⁷) may be quite different depending on whether star formation is primarily regulated by internal properties, such as gas surface density in a quiescent

disk, or triggered by an external event such as an interaction. Because the largest samples of LBGs are primarily identified in the rest UV, model predictions are also quite sensitive to the high-stellar-mass slope of the IMF, and to dust extinction. At the other end of the spectrum is the sub-mm population detected by SCUBA, believed to be predominantly high redshift ($z \gtrsim 2$) luminous and ultraluminous infrared galaxies (LIRGs and ULIRGs) powered by star formation rates of hundreds to thousands of solar masses per year (e.g., 18). Theoretical predictions of the numbers and nature of these objects are highly sensitive to the same issues (the dominant mode of star formation, dust, the IMF), but provide a crucial counter-balance to the optical observations. However, the current mismatch between the sensitivity and spatial resolution of optical and sub-mm instrumentation has made it difficult to establish the connection between the two populations observationally.

The Milky Way, like most nearby galaxies, emits the majority of its light in optical and near-IR wavelengths; only about 30% of the bolometric luminosity locally is released in the far-infrared ¹⁹). This was generally believed to be typical of most of the starlight at all redshifts until the discovery of the far-IR part of the EBL by the DIRBE and FIRAS instruments on the COBE satellite, at a level ten times higher than the no-evolution predictions based on the local luminosity function of IRAS galaxies, and representing twice as much energy as the optical background obtained from counts of resolved galaxies ²⁰). This result suggests that either the dust extinction properties of “normal” galaxies change dramatically with redshift, or a population of heavily extinguished galaxies (perhaps analogous to local LIRGs and ULIRGs) is much more common at high redshift than locally, or both. Some of these galaxies may have already been observed, at 15 μm by ISO ²¹), and at 850 μm by SCUBA ²²).

Guideroni et al. ^{23, 11}) showed that their simplified semi-analytic model could reproduce the multi-wavelength data only if they introduced a population of heavily extinguished galaxies with high star formation rates, and with strong evolution of number density with redshift. This population was introduced ad-hoc by ^{23, 11}), but as discussed by these authors, by ²⁴) (based on ²⁵), and also by SPF, the increasing importance of starbursts at high redshift, due to the increasing merger rate and higher gas fractions, is a natural mechanism to produce this population. The models of SPF contain a detailed treatment of mergers and the ensuing collisional starbursts, which

has been calibrated against the merger rate in cosmological N -body simulations ²⁶⁾ and the starburst efficiency in hydrodynamical simulations ^{27, 28)}. Moreover, they produced good agreement with observations of LBGs (e.g. ²⁹⁾) and damped Lyman- α systems (SPF and ³⁰⁾) as well as low redshift galaxies (SP). Therefore, it will be extremely interesting to see if these same models, when combined with the more sophisticated treatment of dust extinction and emission developed by Devriendt, Guiderdoni, and collaborators, will be able to simultaneously reproduce observations over the broad range of wavelengths and redshifts discussed above.

In the next section we briefly describe the ingredients of our models, and then present the results of the predicted EBL. Section 4 presents the implications for γ -ray attenuation, and §5 briefly discusses some alternative treatments and our own conclusions. The work summarized here is a brief, preliminary sample of the results which will soon be presented in a series of papers, now in preparation, on the EBL and its breakdown into various kinds of sources and on the implications for γ -ray astronomy.

2 Semi-analytic modelling

In this section we briefly describe the ingredients of our models. Readers can refer to SP and SPF for more details, and to ³³⁾ for a brief introduction.

Using the method described in ¹³⁾, we create Monte-Carlo realizations of the masses of progenitor halos and the redshifts at which they merge to form a larger halo. These “merger trees” reflect the collapse and merger of dark matter halos within a specific cosmology (each branching in the tree represents a halo merging event — for examples, see e.g. ³⁴⁾). We truncate the trees at halos with a minimum circular velocity of 40 km/s, below which we assume that the gas is prevented from collapsing and cooling by photoionization. Each halo at the top level of the hierarchy is assumed to be filled with hot gas, which cools radiatively and collapses to form a gaseous disk. The cooling rate is calculated from the density, metallicity, and temperature of the gas. Cold gas is turned into stars using several simple recipes, depending on the mass of cold gas present and the dynamical time of the disk. Supernovae inject energy into the cold gas and may expell it from the disk and/or halo if this energy is larger than the escape velocity of the system. Chemical evolution is traced assuming a constant yield of metals per unit mass of new stars formed. Metals

are initially deposited into the cold gas, and may later be redistributed by supernovae feedback, and mixed with the hot gas or the diffuse (extra-halo) inter-galactic medium.

When halos merge, the galaxies contained in each progenitor halo retain their separate identities either until they spiral to the center of the halo due to dynamical friction and merge with the central galaxy, or until they experience a binding merger with another satellite galaxy orbiting within the same halo. We take into account subhalo truncation due to tidal effects in the larger halo. All newly cooled gas is assumed to initially collapse to form a disk, and major (nearly equal mass) mergers result in the formation of a spheroid. New gas accretion and star formation may later form a new disk, resulting in a variety of bulge-to-disk ratios at late times.

For an assumed IMF, the stellar SED of each galaxy is then obtained using stellar population models. Here we use the multi-metallicity stellar SEDs of ¹²⁾ for the Salpeter and Kennicutt IMF cases, and the solar metallicity GISSEL models ³⁸⁾ for the Scalo IMF. (We have found that using evolving metallicity rather than solar metallicity SEDs has a relatively small impact on the resulting EBL.) Dust extinction is modelled using an approach similar to that of ¹¹⁾. The optical depth of the disk is assumed to be proportional to the column density of metals. We then use a simple slab geometry where stars and gas are homogeneously mixed, and assign a random inclination to each galaxy to compute the absorption. We use a metallicity dependent extinction curve, following ^{23, 11)}.

All absorbed light is re-radiated at longer wavelength. The galactic dust emission spectrum is represented by a combination of three components: 1) hot dust (as in H_{II} regions), 2) warm dust (as in the diffuse H_I), and 3) cold dust (as in molecular clouds). In the models of Devriendt et al. ¹²⁾, these components are modelled as a mixture of polycyclic aromatic hydrocarbon molecules (PAH), very small grains, and big grains. Big grains may be either cold (~ 17 K), or heated by radiation from star-forming regions (as suggested by observations of typical local starburst galaxies like M82). A set of template spectra is then constructed for galaxies of varying IR luminosity, with admixtures of the various components selected in order to reproduce the observed relations between IR/sub-mm color and IR luminosity. A similar approach was used by ³⁹⁾, using a mixture of a typical Orion-like H_{II} spectrum and an H_I spectrum

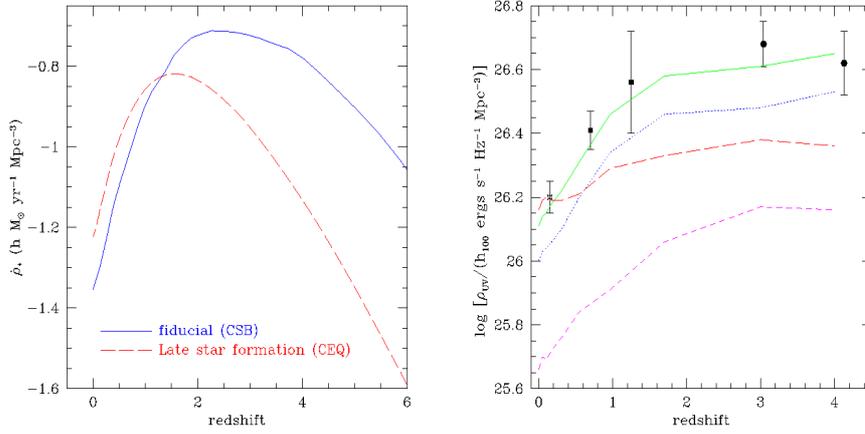


Figure 1: (a) The star formation rate density predicted by our models, for two different recipes of star formation. Both models produce about the same total mass density of stars by $z = 0$ (i.e., the areas under the curves are equal when they are plotted linearly vs. time), but the collisional starburst model (CSB) peaks at higher redshift. (b) Comoving luminosity density at 2000\AA as a function of redshift. Data points represent the observed global luminosity density at rest $\sim 2000\text{\AA}$, obtained by integrating the observational best-fit Schechter luminosity functions ($\rho_L = \phi_* L_* \Gamma(2 - \alpha)$), including corrections for dust extinction. The $z = 0.15$ point is from ³¹⁾, the $z \sim 0.4$ and 1.2 points are from ³²⁾, and the $z \sim 3$ and $z \sim 4$ points are from ¹⁷⁾. The curves for our four models are labeled as in Figure 3. The model curves have been corrected for dust extinction using the approach described in the text.

constructed to fit DIRBE observations of the diffuse ISM ⁴⁰). Here, we use the more empirical emission templates of ³⁹) (kindly provided in electronic form by E. Dwek), but we obtain very similar results with the models of ¹²).

The recipes for star formation, feedback, chemical evolution, and dust optical depth contain free parameters, which we set for each model (see SP) by requiring an average fiducial “Milky Way” galaxy to have a K-band magnitude, cold gas mass, metallicity, and average B-band extinction as dictated by observations of nearby galaxies.

Figure 1a shows the global star formation rate density for the two star formation recipes that we consider here. The “fiducial” model is the collisional starburst (CSB) model favored by SPF, in which bursts of star formation may be triggered by galaxy collisions. The “Late Star Formation” model is the Constant Efficiency Quiescent (CEQ) model of SPF, in which cold gas is converted to stars only in a quiescent mode with constant efficiency. This produces a star formation history similar to the models of the Durham group ⁴¹), in which the

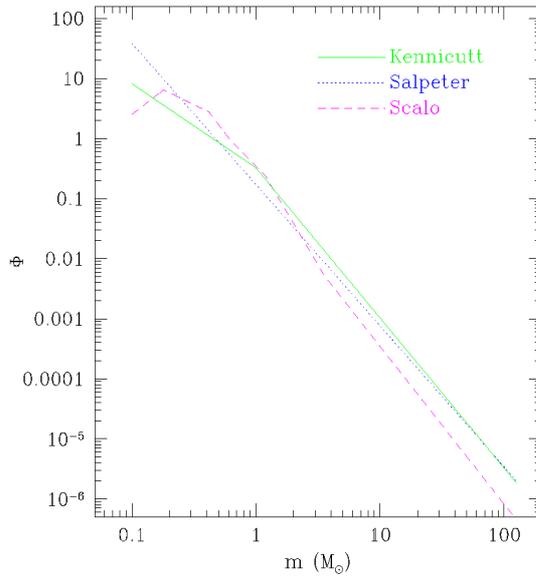


Figure 2: The three stellar Initial Mass Functions (IMFs) used here: Kennicutt ⁴⁴), Salpeter ⁴³), and Scalo ⁴²).

peak in the star formation history occurs at a more recent epoch ($z \sim 1.5$) than in the CSB model. For the CSB model, we consider three different choices of IMF: Scalo ⁴²⁾, Salpeter ⁴³⁾, and Kennicutt ⁴⁴⁾. These IMFs are graphed in Figure 2. For the CEQ model we show only the Kennicutt case. There is a noticeable difference in the far-UV and the mid- to far-IR. The Scalo IMF produces less UV light relative to optical and near-IR light, compared to the Kennicutt and Salpeter IMFs, which produce more high mass stars than the Scalo IMF, and thus more ultraviolet light to be absorbed and re-radiated by dust in the far IR. In Fig. 1b we show the redshift evolution of the far-UV (2000Å) luminosity density for these four models, compared with observations.² The Scalo model falls short at all redshifts, and the CEQ model, which agrees at $z = 0$, falls short at higher redshifts.³ It is encouraging that our very simple model for dust extinction, which we normalized in the B-band at $z = 0$, appears to yield the appropriate level of dust extinction in the UV at higher redshifts (SPF).

Recently, improved luminosity functions (LFs) have become available in optical bands from the Sloan Digital Sky Survey ⁴⁷⁾ and in the K-band from 2MASS ^{48, 49)}. We have found that our CSB model with Kennicutt IMF agrees well with all of these LFs when the average baryon fraction in galaxies is $f_b = 0.1$. This model is also consistent with the number counts in the mid-IR (15 μm from ISOCAM ²¹⁾) and far IR (60 μm from IRAS, 175 μm from ISOPHOT) but not the sub-mm (850 μm from SCUBA ²²⁾). The resulting EBL is similar to that from the Salpeter model discussed below.

²These models were also compared with the observed luminosity density from nearby galaxies, obtained by integrating the luminosity functions of galaxies resolved in recent redshift surveys at wavelengths ranging from 0.2 to 2.2 μm , in Fig. 3 of ¹⁾. However, the SAM outputs graphed there were inadvertently multiplied by a factor of $h^4 \approx 0.25$. In ³³⁾, we renormalized all the models by requiring that they all agreed with the K-band point at 2.2 μm . Here we do not do this since our current SAMs ^{9, 10)} use a corrected version ⁴⁵⁾ of the Press-Schechter formalism.

³The Durham type models ⁴¹⁾ of SPF also predict that LBGs have higher stellar mass than observations indicate ²⁹⁾, while predicted stellar masses from the CSB model of SPF are in good agreement with the observations ⁴⁶⁾.

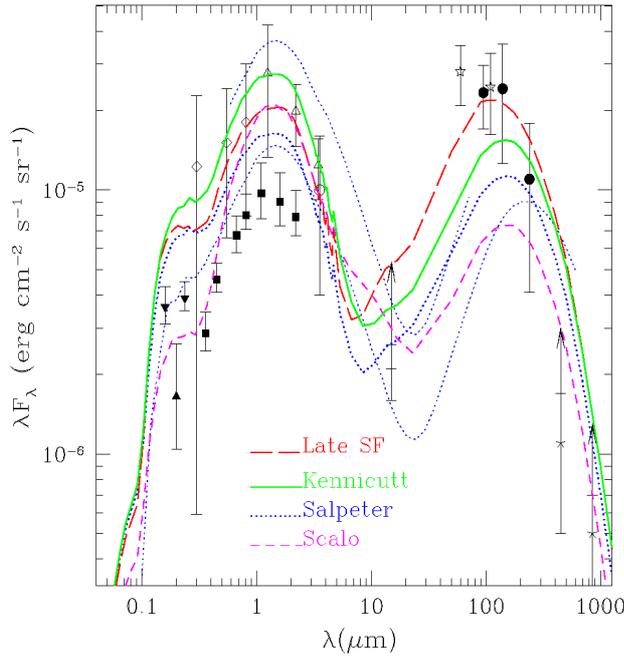


Figure 3: Extragalactic background light: models and data. The far-UV points are from STIS (inverted filled triangles) ⁵⁰) and FOCA observations (filled triangle) ⁵¹). The lower optical points (filled squares) are lower limits from resolved sources ²⁰); the upper ones (open diamonds) are from absolute photometry ⁵²). The near-IR points are from DIRBE: (open circle) ⁵³), (open triangles) ⁵⁴). The point at $15 \mu\text{m}$ is from ISOCAM resolved sources ²¹), and is thus a lower limit. The far-IR points are from DIRBE (filled circles) ^{5, 55}), (stars) ⁵⁶). The curves are our results from modelling the history of star formation in the ΛCDM cosmology using semi-analytic methods: a model with both quiescent star formation with constant efficiency and starbursts, with Kennicutt, Salpeter, and Scalo IMFs, and a Late SF model with only quiescent star formation with constant efficiency (CEQ). The lower light dotted curve is the ΛCDM EBL calculated using our previous methods ³³) for the Salpeter IMF, and the upper one is the same curve to $80 \mu\text{m}$ multiplied by 2.5 for comparison with Mrk 501 data as analyzed by ⁶²) (see text). Note that $10^{-6} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ sr}^{-1} = 1 \text{ nW m}^{-2} \text{ sr}^{-1}$.

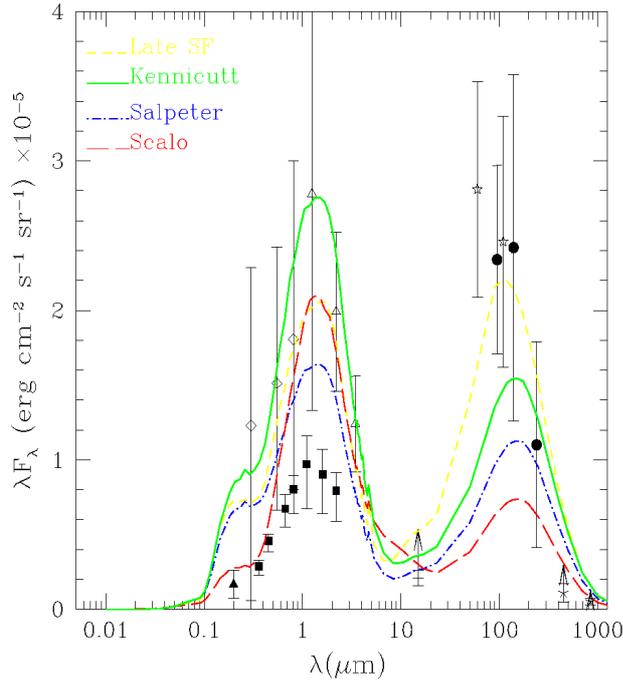


Figure 4: Extragalactic background light: models and data with linear vertical axis. Labels are as in Fig. 3.

3 The Integrated Extragalactic Background Light

Figure 3 shows the EBL produced by our four models, obtained by integrating the light over redshift (out to $z = 4$) with the appropriate K-corrections due to cosmological redshifting. We compare this with a compilation of observational limits and measurements of the EBL. Fig. 4 presents the same four models and the same data, but with a linear rather than logarithmic vertical axis so that one can integrate the total energy in the EBL by eye. It is apparent that there is at least as much energy in the far-IR part of the EBL as in the entire optical and near-IR bands. For example, Puget and collaborators²⁾ estimated that the total energy in the EBL is between 60 and $93 \text{ nW m}^{-2} \text{ sr}^{-1}$, with between 20 and $41 \text{ nW m}^{-2} \text{ sr}^{-1}$ contributed by the optical and near-IR, and between 40 and $52 \text{ nW m}^{-2} \text{ sr}^{-1}$ coming from the far-IR. If the possible detection of

the EBL at $60 \mu\text{m}$ by Finkbeiner et al. ⁵⁶⁾ were correct, that would further increase the far-IR EBL; however, it is very difficult to determine the EBL at $60 \mu\text{m}$ since the zodiacal light is so much brighter at that wavelength, and it was probably partly confused with the EBL ⁵⁷⁾.

The total energy in the EBL in units of critical density ρ_c is $\Omega_{\text{EBL}} = (4\pi/c)(I_{\text{EBL}}/\rho_c c^2) = 2.5 \times 10^{-8} I_{\text{EBL}} h^{-2}$, where I_{EBL} is in units of $\text{nW m}^{-2} \text{sr}^{-1}$. The total energy density in the EBL corresponding to the lower and upper estimates of ²⁾ is $\Omega_{\text{EBL}} = (3.6 - 5.5) \times 10^{-6} (h/0.65)^{-2}$. Although the EBL includes energy radiated by active galactic nuclei (AGNs) as well as stars, it is unlikely that AGNs contributed more than a few percent of the total. This is because the total energy radiated by AGNs is $E_{\text{EBL}}^{\text{AGN}} = \eta \rho_{\text{BH}} c^2$, where the efficiency of conversion of mass to radiated energy in AGNs is $\eta \sim 0.05$. Correspondingly, $\Omega_{\text{EBL}}^{\text{AGN}} = \eta \Omega_{\text{BH}} (1 + z_{\text{BH}})^{-1} \approx 4.5 \times 10^{-8} h^{-1} (\eta/0.05) [3/(1 + z_{\text{BH}})] \lesssim 0.02 \Omega_{\text{EBL}}$. ⁴ So for simplicity, in this paper we will neglect the contribution of AGNs to the EBL.

Several interesting features emerge from the comparison of our SAM models with the EBL data. In the UV to near-IR, the models are closer to the direct measures of the EBL obtained by ^{52, 53, 54)} than to the lower limits from the Hubble Deep Field ²⁰⁾, although the Salpeter IMF produces less light in the UV because it has fewer high-mass stars. As noted, our Kennicutt CSB SAM with $f_b = 0.1$, which agrees well with the latest observed local luminosity density at $z = 0$, produces an EBL close to the Salpeter one in Fig. 4. Of our four new EBL curves, the Late SF model and the fiducial Kennicutt model are also consistent with the DIRBE measurements at $140 \mu\text{m}$. The Salpeter EBL lies a little more than 2σ below the DIRBE measurement at $140 \mu\text{m}$. The LateSF far IR is higher than the other models because its later star formation suffers less dilution due to the expansion of the universe. The models differ significantly in the mid-IR, $\sim 10 - 60 \mu\text{m}$, where the EBL can be probed by TeV γ -rays. The lower dotted curve in Fig. 3, representing our previous attempt ³³⁾ to

⁴Updating ⁵⁸⁾, we have estimated $\Omega_{\text{BH}} = (M_{\text{BH}}/M_{\text{spheroid}})\Omega_{\text{spheroid}} \approx (1.5 \times 10^{-3})(1.8 \times 10^{-3} h^{-1})$, using the observed (loose) correlation ⁵⁹⁾ between a black hole mass and that of the galactic spheroid in which it is found, and the estimated cosmological density of spheroids ⁶⁰⁾. Note that the factor $(1+z)^{-1}$ arises because of the dilution of the contribution of high-redshift sources due to the expansion of the universe.

model the EBL, is well below the $15 \mu\text{m}$ lower limit as well as the DIRBE measurements at longer wavelengths. As we stated in 33), we expected our EBL results to change as we improved our dust emission modelling. In addition to inclusion of the PAH features, the new dust emission model has more warm dust than the one used in 33).

We now discuss constraints from the TeV γ -ray observations.

4 Attenuation of high-energy γ -rays

Figure 5 shows the γ -ray attenuation predicted by the four Λ CDM models considered here, for sources at redshifts $z_s = 0.03$ and 0.10 . All of the models predict rather little absorption at $E_\gamma \lesssim 5 \text{ TeV}$ for sources at $z_s = 0.03$, but

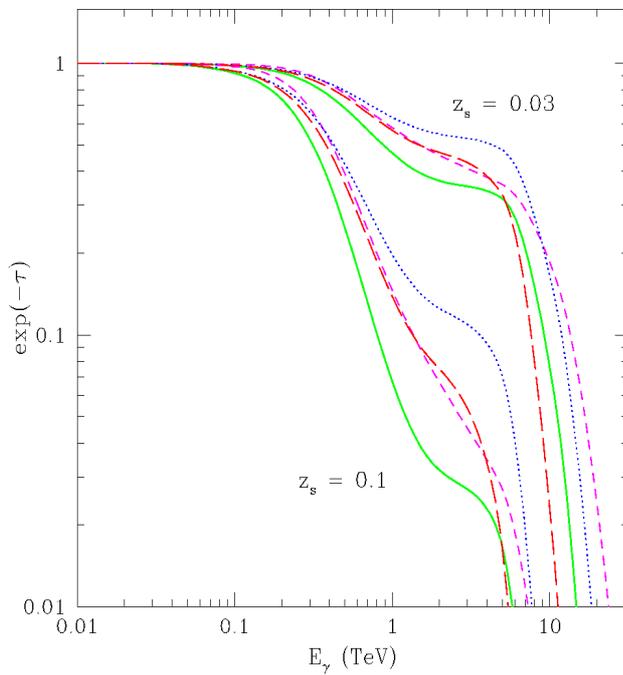


Figure 5: The attenuation factor, $\exp(-\tau)$ for γ -rays as a function of γ -ray energy for the four Λ CDM models considered in Fig. 4. The assumed redshift of the source, z_s , is indicated for each set of curves.

fairly sharp cutoffs above ~ 5 TeV, especially for the Late SF model. That model may be in conflict with the data from Mrk 501. For the blazars Mrk 421 and 501, both at $z \approx 0.03$, the synchrotron self-Compton (SSC) model, in which \sim keV synchrotron X-radiation from a very energetic electron beam is Compton up-scattered by the same electrons to produce the observed \sim TeV γ -rays, appears to explain both the keV-TeV spectra and their time variation (see, e.g., ^{62, 63}) and references therein). Using a simplified SSC model and keV X-ray data to predict the unattenuated TeV spectrum of Mrk 501, Guy et al. ⁶²) used CAT and HEGRA data to estimate the amount of γ -ray attenuation. They find that there is a rather good fit to the observed attenuation for the Λ CDM-Salpeter EBL from our earlier work ³³) when it is scaled upward by a factor of up to about 2.5 across the wavelength range 1-80 μ m; this is the upper Salpeter curve on Fig. 3. Our new Salpeter curve appears to be rather consistent with this rescaling of our old Salpeter one, the Kennicutt curve may be a little high, and the Late SF curve appears to be definitely too high. As noted earlier, the new Salpeter EBL curve in Figs. 3,4 is similar to our latest $f_b = 0.1$ Kennicutt SAM, which is in good agreement with the latest local luminosity functions in the optical and K bands from SDSS and 2MASS, and also in good agreement with IR number counts from IRAS and ISO satellites.

The large flares in Mrk 501 in spring 1999 allowed an accurate measurement of the gamma ray spectrum up to about 17 TeV, and indicated that the spectrum had an exponential cutoff at about 5 TeV ⁶⁴). The flaring activity in Mrk 421 in early 2001 has now provided evidence for an exponential cutoff at about 4 TeV ⁶⁵). The coincidence in the cutoffs (within observational uncertainties) for these two different extragalactic sources suggests that both are due to absorption via pair production on the EBL. In order to confirm this, it will of course be necessary to see similar cutoffs at lower energies for blazars at greater distances. There are already indications of this from 1ES 1426+42, a blazar at redshift $z = 0.13$ (four times farther than Mrk 421), on which there is nearly enough data from CAT ⁶⁶), Whipple, and HEGRA to begin to determine the spectrum from a few hundred GeV to several TeV. It will be very useful to measure the spectrum from this and other sources at comparable distances, such as PKS2155-304 at $z = 0.116$, which will soon be possible with the next generation of atmospheric Cherenkov telescopes such as the H.E.S.S. array in Namibia, the CANGAROO-III array in Australia, and

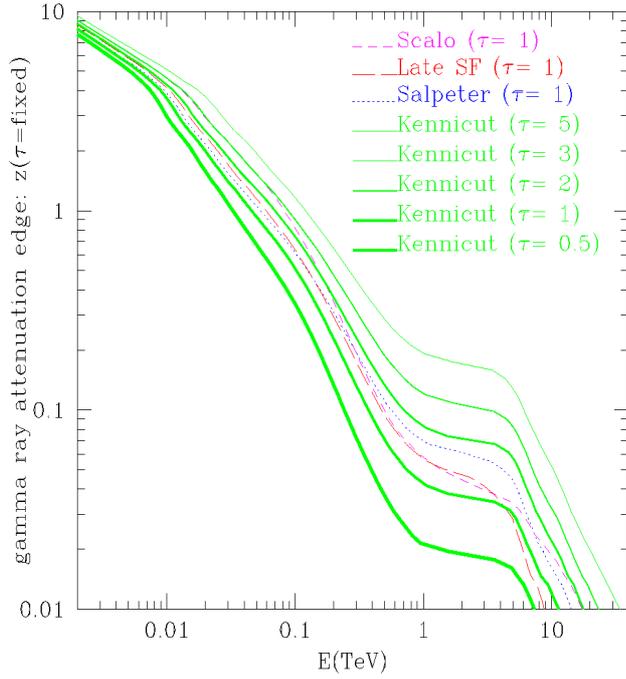


Figure 6: The γ -ray attenuation edge. The redshift where the optical depth reaches unity is shown as a function of γ -ray energy for each of the four Λ CDM models considered in Fig. 4. Also shown for the Kennicutt IMF is the redshift where the optical depth equals 0.5, 2, 3, and 5.

the VERITAS array in Arizona.

Assuming that the EBL is like the LateSF curve in Figs. 3,4, several authors (e.g. ⁶⁷) have argued that the TeV attenuation that this implies plus the observed spectrum of Mrk 501 leads to the requirement that the spectrum at the source have a strong upturn in its emitted flux above about 15 TeV, which would be very hard (although perhaps not impossible ⁶⁹) to understand, and they even suggest that one might have to abandon Lorentz invariance. However, if one instead assumes our Salpeter EBL in Fig. 4 and applies an SSC analysis to the HEGRA data from the Mrk 501 flares in 1997, the implied source spectrum is very reasonable, without an upturn at the high-energy end ⁷⁰.

The compatibility of our new EBL calculations with the available data

on TeV γ -ray attenuation is definitely worth further investigation. The results appear to be sensitive to the details of the models, raising the hope that they may be able to help answer important questions about star formation and dust reradiation, and also help to test the SSC modelling.

Figure 6 depicts the γ -ray “absorption edge,” the redshift of a source corresponding to an optical depth of unity, as a function of γ -ray energy. Travelling through the evolving extragalactic radiation field, γ -rays from sources at lower redshift suffer little attenuation. The universe becomes increasingly transparent as E_γ decreases, probing the background light at increasingly short wavelengths. (We are using the treatment of [61](#)) to account for absorption of ionizing radiation by the Lyman alpha forest.) The models all have the same qualitative features, but differ significantly quantitatively. The location of the absorption edge is affected both by the assumed IMF and by the history of star formation. There is more absorption at most redshifts with the Kennicutt IMF because with a higher fraction of high mass stars, it is more efficient at producing radiation for a given stellar mass; there is more absorption nearby in the Late SF model because the starlight in this model is less diluted by the expansion of the universe. It is possible that measuring the transparency of the universe to γ -rays at ~ 100 GeV with a number of sources at various redshifts can provide a strong probe of star formation, although there are uncertainties due to extinction by dust.

5 Outlook

The semi-analytic modelling of the EBL described here follows the evolution of galaxy formation in time. Forward modelling is a more physical approach than backward modelling (luminosity evolution). Pure luminosity evolution (e.g., [71](#), [72](#), [68](#)) assumes that the entire evolution of the luminosity of the universe arises from galaxies in the local universe just becoming brighter at higher redshift by some power of $(1+z)$ out to some maximum redshift. It effectively assumes that galaxies form at some high redshift and subsequently just evolve in luminosity in a simple way. This is at variance with hierarchical structure formation of the sort predicted by CDM-type models, which appears to be in better agreement with many sorts of observations.

An alternative approach to modelling the EBL has been followed by Pei and collaborators [73](#), [74](#), [75](#), in which they find an overall fit to the global

history of star formation subject to constraints from input data including the evolution of the amount of neutral hydrogen in damped Ly α systems (DLAS). Their first attempt (73, 74), which was used as the basis for EBL estimates by (39, 76), was somewhat misled by the sharp drop in the DLAS hydrogen abundance from redshift $z \sim 3$ to $z \sim 2$ reported in (77). With more complete data on DLAS (see, e.g., Fig. 14 of (78)) the $z = 3$ point is lower and the neutral hydrogen abundance is almost constant from $z = 2$ to 4. The latest paper by Pei et al. (75) takes a variety of recent data into account. Their approach is to follow the evolution of the total mass in stars, interstellar gas, and metals in a representative volume of the universe; they assume a Salpeter IMF. By contrast, the semi-analytic methods we use follow the evolution of many individual galaxies in the hierarchically merging halos of specific CDM models, here Λ CDM. Despite the differences in approach, and the fact that (75) assumed $\Omega_m = 1$ and Hubble parameter $h = 0.5$, their results are broadly similar to those from the semi-analytic approach (see their §4.4). In particular, their EBL is similar to our old results (33) for the Salpeter IMF. Our EBL results presented here are higher in the near-IR and more consistent with the direct determinations (53, 54); they are also higher in the mid-IR, probably mainly because of the warm dust and PAH features in our dust emission model. It will be interesting to see whether further development of the global approach of Pei et al. and of the semi-analytic approach lead to convergent results.

As our calculations show, the EBL, especially at $\lesssim 1 \mu\text{m}$ and $\gtrsim 10 \mu\text{m}$, is significantly affected by the IMF and the absorption of starlight and its reradiation by dust, as well as by the underlying cosmology. The cosmological parameters are becoming increasingly well determined by other observations. As data become available on γ -ray emission and absorption from sources at various redshifts, especially from the new generation of Atmospheric Cherenkov Telescopes and the new γ -ray satellites AGILE and GLAST, these data and their theoretical interpretation will help to answer fundamental questions concerning how and in what environments all the stars in the universe formed.

Acknowledgments

I thank my collaborators Rachel S. Somerville, James S. Bullock, and Julien E. G. Devriendt. My work was supported by NASA and NSF grants at UCSC. I am grateful for a Humboldt Award, and I thank Leo Stodolsky for hospitality

and Eckart Lorenz for enlightening discussions about γ -ray astronomy at the Max-Planck-Institut für Physik, München. I also thank Simon White for hospitality at MPI Astrophysics in Garching, and Heinz Völk for hospitality at MPI Nuclear Physics in Heidelberg. I thank Henric Krawczynski, Paolo Coppi, and Felix Aharonian for very helpful discussions of their SSC modelling. I thank Aldo Morselli for inviting me to give these lectures at the ISSS, and for his patience in waiting for me to send him the written versions.

References

1. J.R. Primack, R.S. Somerville, J.S. Bullock, and J.E.G. Devriendt, in *High Energy Gamma Ray Astronomy*, Proceedings of the International Symposium Gamma-2000, Heidelberg, June 2000, eds. F. Aharonian and H. Völk, AIP Conf. Proc., Vol. 558, pp. 463-478 (2001).
2. R. Gispert, G. Lagache, and J.L. Puget, *A&A* **360**, 1 (2000).
3. J.L. Puget et al., *A&A* **308**, L5 (1996).
4. B. Guiderdoni et al., *Nature* **390**, 257 (1997).
5. M.G. Hauser et al., *ApJ* **508**, 25 (1998).
6. D.J. Fixsen et al., *ApJ* **508**, 123 (1998).
7. G. Kauffmann, S.D.M. White, and B. Guiderdoni, *ApJ* **264**, 201 (1993).
8. S. Cole et al., *MNRAS* **271**, 781 (1994).
9. R.S. Somerville and J.R. Primack, *MNRAS* **310**, 1087 (1999). (SP)
10. R.S. Somerville, J.R. Primack, and S.M. Faber, *MNRAS* **320**, 504 (2001). (SPF)
11. J.E.G. Devriendt and B. Guiderdoni, *A&A* **363**, 851 (2000).
12. J.E.G. Devriendt, B. Guiderdoni, and R. Sadat, *A&A* **350**, 381 (1999).
13. R.S. Somerville and T. Kolatt, *MNRAS* **305**, 1 (1999).
14. R.S. Somerville, G. Lemson, T. Kolatt, and A. Dekel, *MNRAS* **316**, 479 (2000).

15. J.R. Primack, "Cosmological Parameters 2000," in *Sources and Detection of Dark Matter in the Universe*, Proc. 4th International Symposium (DM 2000), Marina del Rey, California, 20-23 Feb 2000, ed. D. Cline (Berlin: Springer, 2001), pp. 3-17 — and also my other lectures in these proceedings.
16. F.R. Pearce et al., *MNRAS* **326**, 649 (2000).
17. C. Steidel et al., *ApJ* **519**, 1 (1999).
18. D.B. Sanders, *Astrophys. and Space Sci.* **269/270**, 381 (1999).
19. B.G. Soifer and G. Neugebauer, *AJ* **101**, 354 (1991).
20. P. Madau and L. Pozzetti, *MNRAS* **312**, L9 (2000).
21. D. Elbaz et al., *A&A Letters* **351**, L37 (1999).
22. A. Blain, I. Smail, R.J. Ivison, and J.-P. Kneib, *MNRAS* **302**, 632 (1999).
23. B. Guiderdoni, E. Hivon, F.R. Bouchet, and B. Maffei, *MNRAS*, **295**, 877 (1998).
24. J. Silk and J. Devriendt, (2000) in in Proc. IAU Symp. 204 *The Extragalactic Infrared Background and its Cosmological Implications*, eds M. Harwit and M.G. Hauser (Astronomical Society of the Pacific, San Francisco, 2001).
25. C. Balland, J. Silk, and R. Schaeffer *ApJ* **497**, 541 (1998).
26. T. Kolatt et al., astro-ph/0010222 (2000).
27. C. Mihos and L. Hernquist, L. *ApJ* **425**, 13 (1994); **448**, 41 (1995); **464**, 641 (1996).
28. R.S. Somerville et al., in XVth IAP Meeting, Dynamics of Galaxies ed. F. Combes et al., ASP Conf. Series **197**, 331 (1999).
29. C. Papovich, M. Dickinson, and H.C. Ferguson, *ApJ*, **559**, 620 (2001).
30. A.H. Maller, J.X. Prochaska, R.S. Somerville, and J.R. Primack, *MNRAS* **326**, 1475 (2001).
31. M. Sullivan et al., *MNRAS* **312**, 442 (2000).

32. L.L. Cowie, A. Songaila, and A.J. Barger, *AJ* **118**, 603 (1999).
33. J.R. Primack, J.S. Bullock, R.S. Somerville, and D. MacMinn, *Astroparticle Phys.* **11**, 93 (1999); J.S. Bullock, R.S. Somerville, D. MacMinn, and J.R. Primack, *Astroparticle Phys.* **11**, 111 (1999).
34. R.H. Wechsler, J.S. Bullock, J.R. Primack, A.V. Kravtsov, and A. Dekel, *ApJ* in press, astro-ph/0108151 (2001).
35. S.R. Folkers et al., *MNRAS* **308**, 459 (1999).
36. M.J. Geller et al., *AJ* **114**, 2205 (1997).
37. J.P. Gardner, R.M. Sharples, C.S. Frenk, and B.E. Carrasco, *ApJ* **480**, L99 (1997).
38. G. Bruzual and S. Charlot, *ApJ*, **405**, 538 (1993).
39. E. Dwek et al., *ApJ* **508**, 106 (1998).
40. E. Dwek et al., *ApJ* **484**, 779 (1997).
41. C.M. Baugh, S. Cole, C.S. Frenk, and C.G. Lacey, *ApJ* **498**, 504 (1998).
42. J.M. Scalzo *Fund. Cosmic Phys.* **11**, 1 (1986).
43. E.E. Salpeter, *ApJ* **121**, 161 (1955).
44. R.E. Kennicutt, *ApJ* **272**, 54 (1983).
45. R. Sheth, and G. Tormen, *MNRAS* **308**, 119 (1999).
46. J.R. Primack, in *Proc. Galaxy Masses at Low and High Redshift* (Venice October 2001), eds. R. Bender and A. Renzini (Springer-Verlag, 2002), in prep.
47. M.R. Blanton et al., *AJ* **121**, 235 (2001).
48. C.S. Kochanek et al., *ApJ* **560**, 566 (2001).
49. S. Cole et al., *MNRAS* **326**, 255 (2001).
50. J.P. Gardner, T.B. Brown, and H.C. Ferguson, *ApJ* **542**, L79 (2000).

51. C. Armand, B. Milliard, and J.-M. Deharveng, *A&A* **284**, 12 (1994).
52. R.A. Bernstein, W.L. Freedman, and B.F. Madore, *ApJ* submitted (2001).
53. E. Dwek and R.G. Arendt, *ApJ* **508**, L9 (1998).
54. V. Gorjian, E.L. Wright, and R.R. Chary, *ApJ* **536**, 550 (2000); E.L. Wright, astro-ph/0004192 (2000).
55. G. Lagache et al. in *ISO Surveys of a Dusty Universe*, ed. D. Lemke et al., Springer Lecture Notes in Physics **548**, 81 (2000).
56. D.P. Finkbeiner, M. Davis, and D. Schlegel, *ApJ* **544**, 81 (2000).
57. D.P. Finkbeiner, in Proc. IAU Symp. 204 *The Extragalactic Background and its Cosmological Implications*, eds. M. Harwit and M.G. Hauser (Astronomical Society of the Pacific, San Francisco, 2001), p. 121.
58. P. Madau, *VLT Opening Symposium* ed. J. Bergeron et al. (Springer, Berlin, 2000), p. 52.
59. J. Kormendy, astro-ph/0007401 (2000).
60. M. Fukugita, C. Hogan, and P.J.E. Peebles, *ApJ* **503**, 518 (1998).
61. P. Madau, *ApJ*, **441**, 18 (1995).
62. J. Guy, C. Renault, F.A. Aharonian, M. Rivoal, and J.-P. Tavernet, *A&A* **359**, 419 (2000).
63. H. Krawczynski, P.S. Coppi, T. Maccarone, and F.A. Aharonian, *A&A* **353**, 97 (1999).
64. F. Aharonian et al., *ApJ* **546**, 898 (2001).
65. F. Krennrich et al., *ApJ* **560**, 45 (2001).
66. A. Djannati-Atai et al., in *Proc. ICRC-2001*.
67. R.J. Protheroe and H. Meyer, *Phys. Lett.* **B493**, 1 (2000).
68. F. Stecker, in Proc. IAU Symp. 204 *The Extragalactic Background and its Cosmological Implications*, eds. M. Harwit and M.G. Hauser (Astronomical Society of the Pacific, San Francisco, 2001).

69. F. Aharonian, A.N. Timokhin, and A.V. Plyasheshnikov, *A&A* submitted, astro-ph/0108419 (2001).
70. H. Krawczynski, P. Coppi, and F. Aharonian, submitted (2001).
71. M.A. Malkan and F.W. Stecker, *ApJ* **496**, 13 (1998).
72. M.A. Malkan and F.W. Stecker, *ApJ* **555**, 641 (2001).
73. Y.C. Pei and S.M. Fall, *ApJ* **454**, 69 (1995).
74. S.M. Fall, S. Charlot, and Y.C. Pei, *ApJ* **464**, L43 (1996).
75. Y.C. Pei, S.M. Fall, and M.B. Hauser, *ApJ* **522**, 604 (1999).
76. M.W. Salamon and F.W. Stecker, *ApJ* **493**, 547 (1998).
77. K.M. Lanzetta, A.M. Wolfe, and D.A. Turnshek, *ApJ* **440**, 435 (1995).
78. L. Storrie-Lombardi and A.M. Wolfe, *ApJ* **545**, 603 (2000).

Participants

AMELINO CAMELIA Giovanni	Univ. Roma “La Sapienza”
ARAMO Carla	INFN – Catania
ARDIZZOIA Gabriella	CIFS - Torino
AURICCHIO Natalia	Ist. TESRE-CNR – Bologna
BALBI A.	Univ. Roma “Tor Vergata”
BARBIELLINI Guido	Universita' di Trieste
BATTISTON Roberto	Universita' di Perugia
BELLAZZINI Ronaldo	Universita' di Pisa
BENCIVENNI Giovanni	INFN – Lab. Nazionali Frascati (RM)
BEREZHKO Evgeny	Inst. Cosmophysical Research – Yakutsk – Ru
BEREZINSKY Venya	LNGS, INFN – Assergi
BIANCHI Stefano	Università Roma3
BOEZIO Mirko	INFN Trieste
BONGI Massimo	INFN - Univ. Firenze
BOSCHERINI Massimo	University of Siegen
BUTT Yousaf	Harvard-Smithsonian Center for Astrof.
CAMARRI Paolo	Univ. Roma 2 “Tor Vergata”
CARAVEO Patrizia	IFC/CNR – Milano
CASOLINO Marco	Univ. Roma “Tor Vergata”
CELOTTI Annalisa	Sissa – Trieste
CESARINI Alessandro	Univ. Roma Tor Vergata
CIOCCA Claudia	Ist. TESRE-CNR – Bologna
CIPRINI Stefano	INFN – Perugia
COCCO Veronica	INFN – Roma Tor Vergata
COSTA Enrico	CNR/IAS – Roma
DAR Arnon	Technion, Israel Inst. of Technology – Haifa
DE ANGELIS Alessandro	Universita' di Udine
DE BERNARDIS Paolo	Universita' di Roma “La Sapienza”
DE LUCA Andrea	CNR-IFC – Milano
DE PASQUALE Massimiliano	Ist. Astrofisica Spaziale – “La Sapienza”
DELGADO MENDEZ Carlos Jose	CIEMAT – Madrid
DI GIROLAMO Tristano	Dip. Fisica - Università Roma 3

DONATO Fiorenza	Lapth (Francia) – INFN (Roma)
ELLIS JOHN	CERN – Ginevra
ESPOSITO Gennaro	Universita' e INFN – Perugia
FIANDRINI Emanuele	Universita' e INFN – Perugia
FLEURY Patrick	Ecôle Polytechnique – Palaiseau – F
FRATINI Katia	Dip. Fisica – Universita' di Roma 3
FUCITO Francesco	INFN – Roma 2
FURANO Gianluca	INFN- Roma 2
GROSSI Marco	Inghilterra
GUIDORZI Cristiano	Dip. Fisica – Ferrara
IANNUCCI Alessandro	INFN – Univ. Roma
IONICA Maria	INFN – Perugia
KANBACH Gottfried	Max Plank Intitut – Monaco
LANCIOTTI Elisa	CIEMAT – Madrid
LIONETTO Andrea	Università Roma 2 (CIFS)
LONGAIR MALCOM S.	Canendish Laboratory – Cambridge
LONGO Francesco	Università e INFN di Ferrara
MARTANA Simona	Roma2
MENICUCCI Alessandra	Università Roma 2 – INFN
MINORI Mauro	INFN - Roma 2 (CIFS)
MORSELLI Aldo	INFN – Roma 2
NAJAFI Abouzar	Radio Astronomical Instit. – Uni. Bonn
OMODEI Nicola	Università di Siena – INFN
OSTORERO Luisa	Osservatorio Astronomico – Torino
PALAZZO Giusimelissa	Universita' di Catania
PERCOSSI Germano	INFN – Roma
PICOZZA Piergiorgio	Universita' di Roma “Tor Vergata”
PITTORI Carlotta	INFN – Roma Tor Vergata
PREGER Barbara	CIFS-IAS Roma
PRIMACK Joel	University of California – Santa Cruz
REBECCHI Sonia	CIFS – Roma
RICCI Marco	INFN – Lab. Nazionali Frascati (RM)

RICCIARINI Sergio Bruno	INFN - Univ. Firenze
SBARRA Cristina	INFN – Bologna
SCARSI Livio	CNR – Palermo
SETTI Giancarlo	Universita' di Bologna
SIMON Manfred	University of Siegen – D
SOLINI Paola	Int. School of Space Science – L’Aquila
SPARVOLI Roberta	Universita' di Roma “Tor Vergata”
SPILLANTINI Piero	Universita' di Firenze
TAVANI Marco	CNR/IFC – Milano
TOSTI Gino	Dip. Fisica e INFN – Universita' di Perugia
ULLIO Piero	California Inst. of Technology – Pasadena
VIETRI Mario	Univ. Roma3
VILLANTE Umberto	Univ. of L’Aquila – Director of the School
VOLONTE’ Sergio	ESA - Paris
WOZNA Agnieska	N. Copernicus Astronomical Center – PL