

**Lecture Notes and Essays in Astrophysics**  
**VOLUME III**

December 2008

Granada, Spain

Editors

A. Ulla & M. Manteiga

Under the auspices of the Spanish Royal Physical Society

Sponsored by the Spanish Ministry of Science and Innovation

Partially financed by FEDER funds (AYA2007-28639-E).



# Contents

<b>1 THE ATMOSPHERE OF VENUS: WINDS AND CLOUDS OBSERVED BY VIRTIS/VENUS EXPRESS</b>	<b>1</b>
<i>R. Hueso</i>	
<b>2 THE UPPER ATMOSPHERE OF VENUS OBSERVED BY VENUS EXPRESS</b>	<b>13</b>
<i>Miguel Angel López-Valverde et al.</i>	
<b>3 THE PHYSICAL CHARACTERIZATION OF THE STARS</b>	<b>33</b>
<i>C. Jordi, E. Masana</i>	
<b>4 DARK ENERGY AND THERMONUCLEAR SUPERNOVAE</b>	<b>45</b>
<i>I. Domínguez et al.</i>	
<b>5 GALAXIES AND THEIR ENVIRONMENTS: CONNECTIONS WITH NUCLEAR ACTIVITY</b>	<b>63</b>
<i>Isabel Márquez</i>	
<b>6 MAGNETIC FIELDS IN GALAXIES</b>	<b>83</b>
<i>E. Battaner et al.</i>	
<b>7 GRAN TELESCOPIO CANARIAS: A KEY ASSET FOR SPANISH ASTRONOMY</b>	<b>103</b>
<i>F. Sánchez</i>	
<b>8 HIGH-RESOLUTION GROUND-BASED EUROPEAN SOLAR PHYSICS</b>	<b>113</b>

*M. Collados*

**9 ROBOTIC ASTRONOMY AND ITS APPLICATION TO THE STUDY  
OF GAMMA-RAY BURSTS 131**

*A. J. Castro-Tirado*

**10 THE MARTIAN UPPER ATMOSPHERE 151**

*F. González-Galindo et al.*

**11 WHY ARE INTENSE GEOMAGNETIC STORMS SO IMPOR-  
TANT FOR HUMAN LIFE? 163**

*Elena Saiz et al. et al.*

**12 THE PHYSICS OF THE HII REGIONS: THE PHOTOIONIZA-  
TION EQUILIBRIUM IN THE HII REGIONS OF M51 177**

*L. Gutiérrez and J.E. Beckman*

**13 THE FUNDAMENTAL CYCLOTRON LINE IN 4U 1538–52 189**

*José J. Rodes et al.*

**14 IR PROPERTIES OF ASTROPHYSICAL ICES IN THE UNIVERSE 201**

*R. Vilaplana et al.*

**15 AN OVERVIEW OF SOHO 12 YEARS LATER 213**

*R. Gómez-Herrero et al.*

**16 PREPARING ESA Gaia MISSION: STELLAR SPECTRA PARAM-  
ETERIZATION BY NEURAL NETWORKS IN THE NEAR IN-  
FRARED 225**

*D. Ordóñez Blanco et al.*

## FOREWORD

This is the third volume of the series entitled “Lecture Notes and Essays in Astrophysics”, a biannual publication started in 2004 and intending to offer to the specialized community a leading collection of the Spanish Astrophysics research advances, on the basis of a selected sample of lectures presented during the Astrophysics Symposia at the biannual meetings of the Royal Spanish Physical Society (RSEF). In particular, this volume contains the invited reviews (lecture notes) and a selection of the contributions (essays) to the III Astrophysics Symposium that took place at the Science Faculty of the University of Granada during the XXXI Scientific meeting of the RSEF, in September 2007.

The book highlights some important contributions of Spanish astrophysicists to Planetology, Solar and Stellar Physics, Extragalactic Astronomy, Cosmology and astronomical instrumentation. After decades without a dedicated mission, Venus is again in fashion. On the one hand, Ricardo Hueso and collaborators, and on the other Miguel Ángel López-Valverde, review ESA Venus Express contribution to the understanding of the atmosphere of the neighbouring planet.

Carme Jordi describes in a comprehensive essay the main observational calibration techniques and methods for the determination of mass, radius, temperature, chemical composition and luminosity of a star. Dying stars are fundamental to understand the nature of dark energy, probably the most fundamental problem in Physics today. Type Ia supernovae have played a fundamental role showing the acceleration of the expansion rate of the Universe a decade ago. Inma Domínguez and collaborators go into detail on how the knowledge of the fundamental physics of thermonuclear supernovae explosions condition their role as astrophysical candles.

The influence of galactic environments in galaxy activity and the properties of magnetic fields in galaxies are reviewed, respectively, by Isabel Márquez and Eduardo Battaner. First light of the Gran Telescopio Canarias (GTC) is the subject of the review by Francisco Sánchez, encourager of such an endeavour which is today a reality. Robotic astronomy is not the future but a fact materialised for several telescopes worldwide, some of them in Spain. Alberto Castro-Tirado describes some of these instruments and their role in the detection and following up of GRBs. And more.

On behalf of the Astrophysics Group of the RSEF, as in previous volumes, it is the hope of the Editors that this book stimulate the interest for Astronomy, even more being 2009 the International Year of Astronomy.

The Editors are indebted to the Spanish Ministry of Science and Innovation for financial support through grant AYA-2007-28639-E and FEDER funds. This book has been edited under the auspices of the Spanish Royal Physical Society (RSEF).

Ana Ulla (President) and Minia Manteiga (Secretary)  
Astrophysics Group, of the Royal Spanish Physical Society (RSEF)



# THE ATMOSPHERE OF VENUS: WINDS AND CLOUDS OBSERVED BY VIRTIS/VENUS EXPRESS

RICARDO HUESO, SANTIAGO PÉREZ-HOYOS, AGUSTÍN SÁNCHEZ-LAVEGA, AND JAVIER PERALTA

*Física Aplicada I, Escuela Superior de Ingeniería, Universidad del País Vasco, Alda. Urquijo s/n, 48013, Bilbao, SPAIN*

**Abstract:** After decades without a dedicated space mission, Venus, sometimes referred as Earth's twin planet, is again the subject of intense exploration. Venus Express (VEX), a mission of the European Space Agency, was launched in November 2005 and arrived to the planet in April 2006 where it is now in orbit. Among its scientific payload, the Visible and Infrared Thermal Imaging Spectrometer (VIRTIS) is an instrument well suited for the study of a plethora of atmospheric processes through its visual and spectral modes. In this work we will review some of the first results obtained with this instrument. The set of observations characterizes the properties of the multiple cloud layers in Venus, the global atmospheric dynamics and its variability as well as the structure of particular meteorological structures like the polar vortex and the nearly ubiquitous systems of gravity waves.

**Keywords:** Venus – Atmosphere dynamics – Meteorology.

## 1 Introduction

Venus is the planet closest to Earth in distance to the Sun, mass, radius, density and chemical composition. Yet, the similarities end here. Venus has a massive atmosphere made up of carbon dioxide with surface pressures of 90 bar, clouds of sulphuric acid and a hot, dry and young surface according to geological standards. Understanding the similarities and differences between Earth and Venus motivated an intense exploration of this planet paused on the last decades in part by the lure of the exploration of Mars, more attractive to the search of life.

Venus Express is an orbiter based on the Mars Express spacecraft and is the first mission dedicated to Venus atmospheric investigations since NASA's Pioneer Venus and the Soviet Vega and Venera programs, both more than a quarter of a century ago. It was launched from Baikonur in Kazakhstan on November 9, 2005 and arrived to

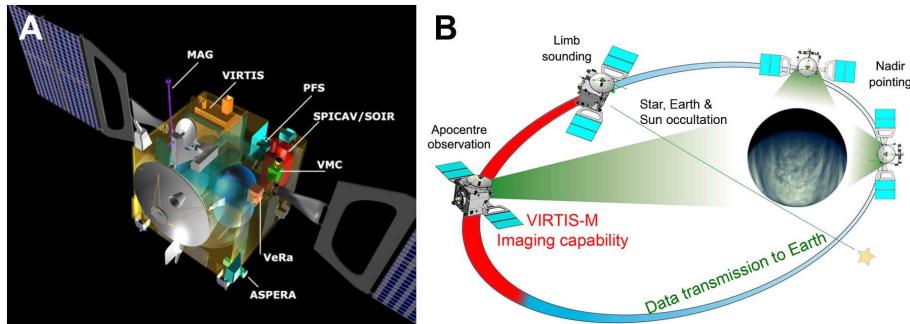


Figure 1: VEX instruments (left) and orbit (right). The scientific payload is composed of ASPERA-4 (Analyser of Space Plasmas and Energetic Atoms), MAG (Magnetometer), PFS (Planetary Fourier Spectrometer; this instrument did not deploy and it is not operational), SPICAV (Spectroscopy for Investigation of Characteristics of the Atmosphere of Venus), VeRa (Venus Radio Science), VIRTIS (Visible and Infrared Thermal Imaging Spectrometer), and VMC (Venus Monitoring Camera). The orbit characteristics allow to obtain nadir observations of Venus South Hemisphere and high-resolution VMC images of Northern latitudes as well as limb-observations including occultation experiments. Data is sent everyday to the Earth in the descending branch of the orbit. The operational maneuvers and data acquisition are complicated by the requirement of keeping the spacecraft and instruments cool under the intense solar radiation while maximizing the data flow to Earth during different observing seasons.

Venus on April 11, 2006. Since then it follows a highly elliptical polar orbit (apocenter at 60,000 km over the South Pole and pericenter at 250 km over the North Pole) with a 24 hr period. The highly elliptical polar orbit combines global nadir observations of the southern hemisphere with close-up snapshots of the equatorial and northern latitudes. The scientific payload consists of seven instruments detailed on Figure 1. While the initial nominal mission was proposed to last for 486 days, roughly 2 Venus days, it has been extended at least to September 2009. The mission description and science operations plans are presented in [1,2].

## 2 Observations

The VIRTIS instrument is an imaging spectrometer inherited from the VIRTIS instrument onboard the Rosetta mission [3]. It is a dual instrument with separate telescopes operating on two channels: (1) VIRTIS-M, a mapping spectrometer with

two CCD detectors working in the visible (VIRTIS-M-vis from 0.3 to 1  $\mu\text{m}$ ) and in the infrared (VIRTIS-M-IR from 1 to 5  $\mu\text{m}$ ) and (2) VIRTIS-H, a high-resolution spectrometer with a spectral range in the infrared 2 – 5  $\mu\text{m}$  ([4,5]). VIRTIS-M obtains image “qubes” using a mirror-scanning technique that simultaneously provides more than 430 images of the same field in the wavelength range of each detector. VIRTIS-M-vis obtains data of the day-side of the planet in reflected light while VIRTIS-M-IR obtains most of its data from thermal radiation escaping from the planet and observed during the night-side (see Figure 2).

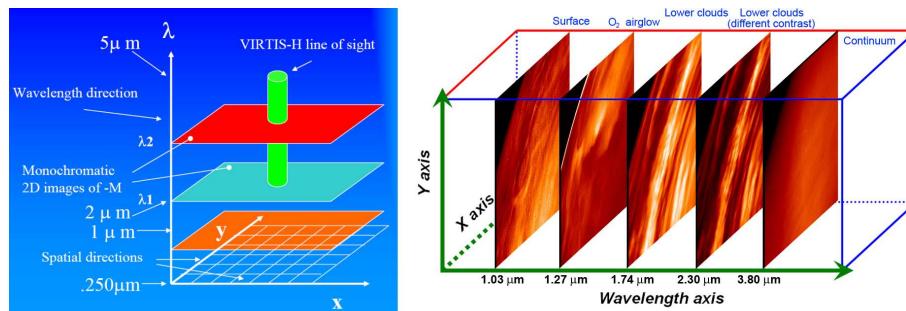


Figure 2: Organization of a VIRTIS data qube. The three-dimensional data qubes contain a full spectra for each x-y pixel and an image for each wavelength. The high-resolution spectral mode allows to obtain 432 images on the spectral range of the instrument (different for VIRTIS-M-vis and VIRTIS-M-IR). The detector is a 256x432 CCD where the spectra produced by a diffraction slit is captured. Each of the 256 pixel lines register a single spectrum. A scanning mirror is used to change the direction of the light obtaining a full data qube in minutes. This information can be used to simultaneously study several vertical layers of the planet, from the surface to the clouds, as well as basic chemical composition and thermal structure of the upper troposphere. The right panel shows representative images of one VIRTIS data qube.

### 3 Venus vertical cloud structure

Venus is known to have a thick and highly reflective cloud cover which fully encircles the planet [6]. This cloud deck is composed of sulphuric acid droplets in a concentration of 75%. When observing Venus reflected spectrum in the vicinity of 1  $\mu\text{m}$  a few notable characteristics arise (see Figure 3). The spectrum is mostly flat from around 500 nm. Below this wavelength an unknown absorber produces a sharp decrease of reflectivity [7]. At even shorter wavelengths (around 300nm) the gaseous

$SO_2$  absorption (highly dependent on different temporal and spatial scales) is also present. Other gaseous absorption is produced by  $CO_2$  above  $1 \mu m$  in various bands of increasing intensity. The  $H_2SO_4$  forming cloud particulates also produce some absorption features closer to  $2 \mu m$  wavelengths. Focusing on the VIRTIS instrument capabilities the information we can retrieve from the analysis of Venus reflected spectrum includes: (a)  $SO_2$  concentration and its variation over time and space; (b) the nature and origin of the UV absorber, together with its temporal variability; (c) the density and particle size of cloud decks; (d) the vertical extent of the cloud layers and its variation with latitude and time. On the other hand, night-side observations in the infrared allow to sound deeper clouds and even the surface through different observation windows as also shown on Figure 3.

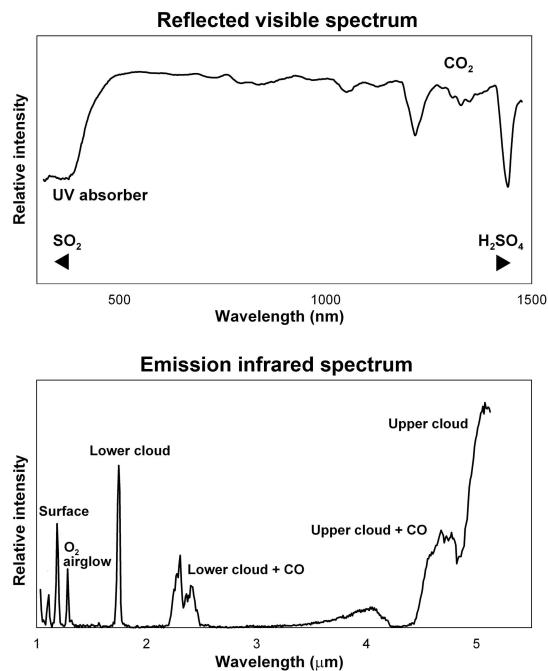


Figure 3: Venus reflected and emitted spectrum. The top panel shows the spectrum in reflected light on the day-side in the vicinity of  $1 \mu m$  with a mean reflective of  $\sim 70\%$ . The bottom panel shows the infrared emission spectrum obtained by VIRTIS on the night-side of the planet with particular observation windows detailed on the figure.

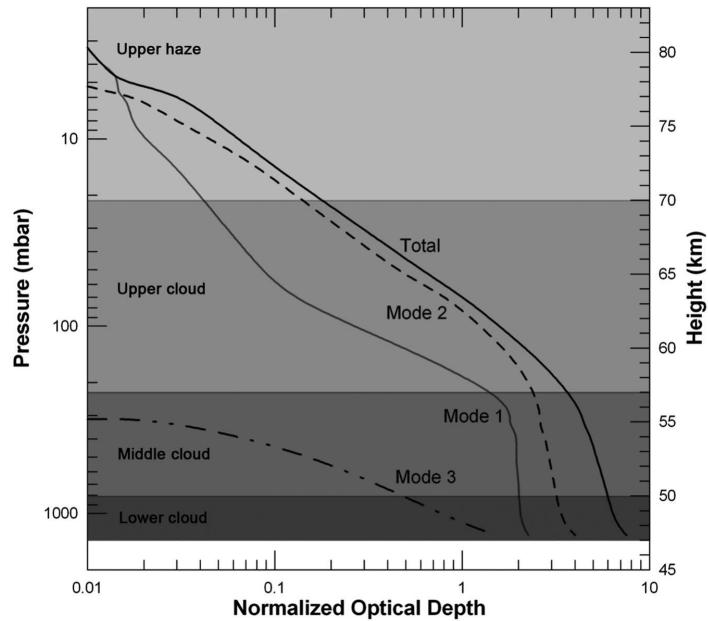


Figure 4: Vertical distribution of particles in Venus atmosphere. Shaded areas mark the extension of the three cloud decks and the haze above. Lines show the contribution to the total optical thickness normalized to the value at 630 nm for an isotropic scatterer. Modes represent different particle sizes. Adapted from [8].

Venus clouds have been analyzed by remote sounding and in situ investigations which were able to characterize the vertical distribution of particles. As summarized on Figure 4 the clouds are accepted to be highly stratified in several extended layers with different types of particles. Most of the visible radiation is reflected from the upper or middle part of the cloud, being the shortest wavelengths strongly affected by the variable presence of the upper haze. Inside each cloud layer, different particle size distributions (“modes”) coexist. Mode 1 corresponds to the smallest particles with radii around  $0.1 \mu\text{m}$ . These are dominant in the upper part of the atmosphere and probably have their origin in the photochemical processes taking place in even higher locations of the atmosphere [6]. At deeper atmospheric levels particles tend to be somewhat larger and Modes 2 (radii  $\sim 1.0 \mu\text{m}$ ) and 3 (radii around  $3.0 - 5.0 \mu\text{m}$ )

are found. The former is the strongest contributing particle size to the total optical thickness of the cloud layers and being representative of the mean particle size. This multilayered cloud structure implies a total optical depth at the surface of  $\sim 30$ .

VIRTIS observations depend strongly on the wavelength and three cloud layers are clearly observed with uncorrelated structures at 380 nm and 980 nm in reflected light and at 1.74  $\mu\text{m}$  in infrared transmitted light. In order to determine the altitude location of these clouds we used the following approach. For 380 and 980 nm we considered two vertical cloud structure models [7,8] including their expected spatial and temporal variability. The total optical depth is calculated as a function of height for the wavelengths of interest [9,10]. For each wavelength the altitude sounded is assumed to be that in which the total optical depth is  $\tau = 1 - 3$ . For observations of transmitted radiation at 1.74  $\mu\text{m}$  the expected sounding level is that at which the maximum concentration of particles is located. With these approaches the sounding levels retrieved are  $\sim 65$  km at 380 nm,  $\sim 60$  km at 980 nm and  $\sim 50$  km, in good agreement with previous estimations [11,12].

## 4 Atmospheric dynamics at cloud levels

### 4.1 Global circulation

The global atmospheric circulation is characterized at cloud level by a zonal super rotation which is one of the most intriguing aspects of the atmosphere [13-18]. Whereas the planet turns around its axis in retrograde sense (from East to West) in 243 days, its atmosphere at the upper cloud level and low latitudes takes only 4 days to encircle the planet. Despite much modeling efforts and observations from different spacecrafts (orbiters, probes and balloons), the underlying mechanism is far from being understood [16].

Venus Express observes the planet in a polar orbit that allows to perform a global study of the South hemisphere of Venus. While the Venus Monitoring Camera (VMC) has provided high-resolution observations of the upper clouds and mean wind profiles at their altitude ( $\sim 65$  km) [19] VIRTIS has been able to study the different cloud layers showing that two fundamentally different circulation regimes separated at latitudes  $55^\circ$  simultaneously operate at the three cloud levels. At low latitudes, zonal winds in the Southern hemisphere at altitudes ranging from 47 to 66 km are nearly constant with latitude with westward velocities of 105 m/s at cloud-tops and 60-70 m/s at the cloud-base. At high latitudes, zonal wind speeds decrease linearly with virtually no vertical wind shear, indicating a vertically coherent vortex structure. Meridional winds at the cloud-tops are poleward with peak speed of 10 m/s at  $55^\circ\text{S}$  but are found to be light below the cloud tops [20]. The peak speed at  $55^\circ\text{S}$  and the rapid wind decrease toward the pole are most probably related to the structure of the polar vortex.

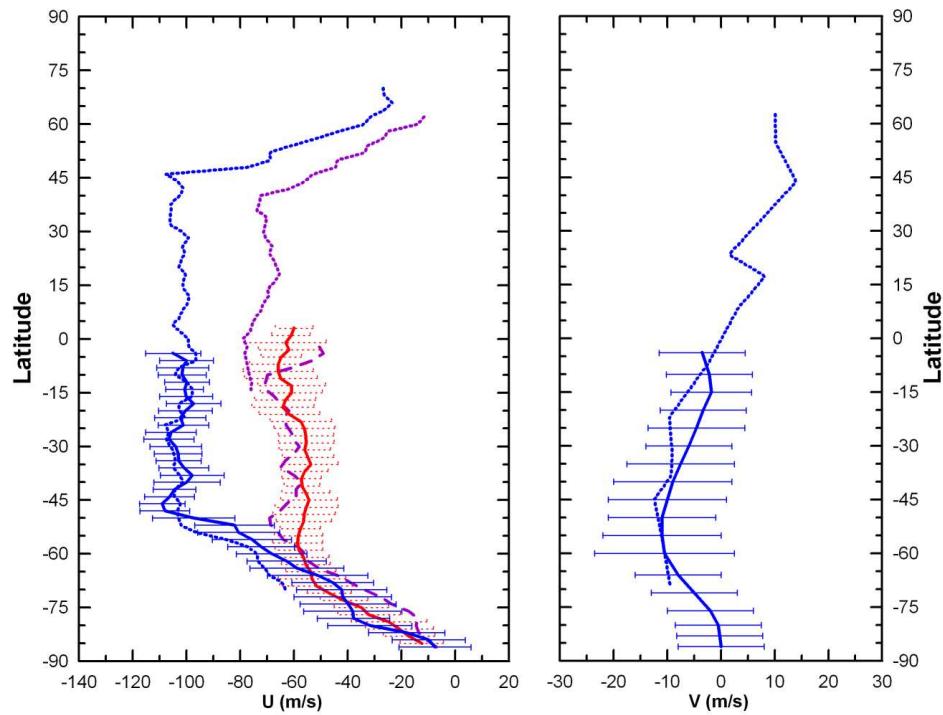


Figure 5: Comparison of zonal and meridional winds obtained by VIRTIS onboard Venus Express in 2007 in the South hemisphere (continuous and dashed lines, [20]) and data from the Galileo spacecraft flyby of Venus in 1991 (dotted lines, [19]). Zonal winds (left panel) are shown with different lines corresponding to winds measured in clouds observed at different wavelengths representative of different vertical altitudes: The left line with error bars represents the winds at 65 km altitude level (observations at 380 nm in the day-side of the planet); the intermediate dashed line without error bars represents the winds at 61 km (observations at 980 nm in the day-side) and the right line with dotted error bars represents the winds at 50 km (observations of the night-side at 1.74 microns). Meridional winds (right panel) indicate a global Hadley cell that transports heat from the Equator to the Poles [20].

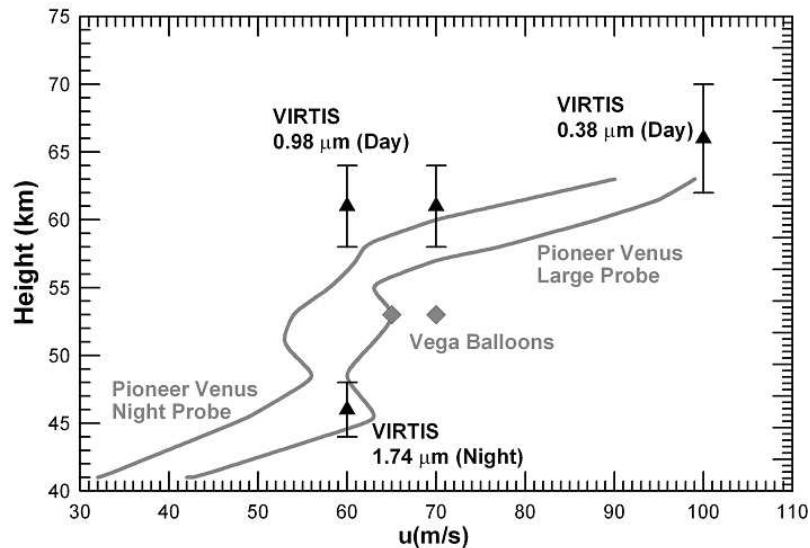


Figure 6: Venus zonal wind speeds as a function of height as measured by Pioneer Venus Probes (lines), Vega Balloons (rhombs) and VIRTIS measurements (triangles with dashed bars). Adapted from [20].

Figure 6 compares our results with the vertical profiles of the zonal winds as measured by Pioneer-Venus entry probes [13,16] in day time (for wavelength 380 nm) and night time (for 1.74  $\mu\text{m}$ ), and with the Vega balloons in night and day time [14]. There is good agreement between the VIRTIS observations and modeling of the cloud altitude and the *in situ* results obtained decades ago.

## 4.2 Winds variability

In spite of remarkable similarities in the wind regimes at low and high-latitudes between different spacecrafts separated decades in time, the winds are known to be variable at the levels of the upper cloud deck. The variability is due to the solar tide and to a 5 days global oscillation not fully understood. Both types of variability had been first discovered during the Galileo flyby of Venus in 1991 [21, 22] at equatorial latitudes but VIRTIS data shows that both oscillations extend to subtropical latitudes (50-70°) where they are even larger in magnitude [20] with wind amplitude variations of 10 m/s.

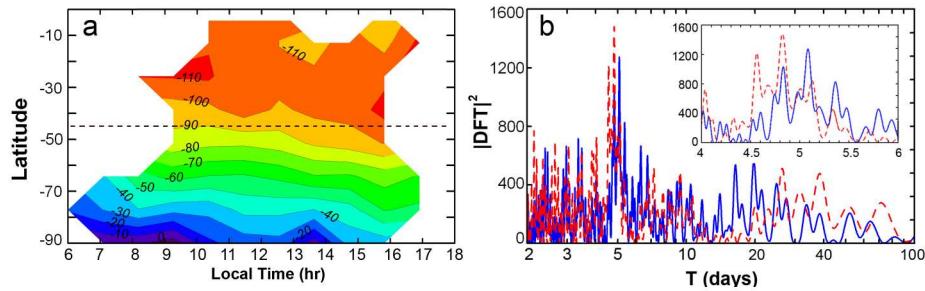


Figure 7: Wind variability in Venus. The left panel is a contour plot with velocities in  $\text{ms}^{-1}$  indicating the dependence of the zonal wind on the local time at each latitude. This effect is specially significant at high to subpolar latitudes. The right picture shows a periodogram of wind values obtained on different days of observations at sub-polar latitudes. The periodogram peaks at 5 days showing a cyclic variation of winds with this period. The two curves correspond to latitude intervals of  $50\text{-}60^\circ$  and  $60\text{-}70^\circ$ .

### 4.3 Vortices, waves and lightning

Among the first results from Venus Express lies the discovery of a double-eye vortex on the South Pole of Venus [23]. The polar dipole is similar to the dipole discovered years ago by different spacecrafts on the North Pole [24]. VIRTIS observations show that the vortex may extend down to the lower cloud layers at 50 km and perhaps deeper. The vortex is 10 K warmer than a surrounding cold collar and it might be an extension of a global Hadley cell descending at polar latitudes. The upper clouds of Venus are also prompt to the development of several wave systems. The most famous is the Y global planetary wave [16] but there is also an intense activity of small-scale wave observed by VEX on the upper and lower clouds [19, 25]. These waves are gravity waves arising on the vertically stable cloud layers [16] and probably excited by convection on the clouds. Their characteristic wavelength is  $\sim 100$  km and their phase speed is  $\sim 10 - 15$  m/s. They are probably gravity waves ducted in the cloud layers where the static stability is high. The waves observed in the upper and lower clouds do not seem to correlate and might be disconnected by the region of low static stability separating both clouds. The convection is assumed to be responsible of the intense lightning activity detected by VEX [26]. The South polar dipole and gravity waves in the upper and lower clouds are illustrated on Figure 8.

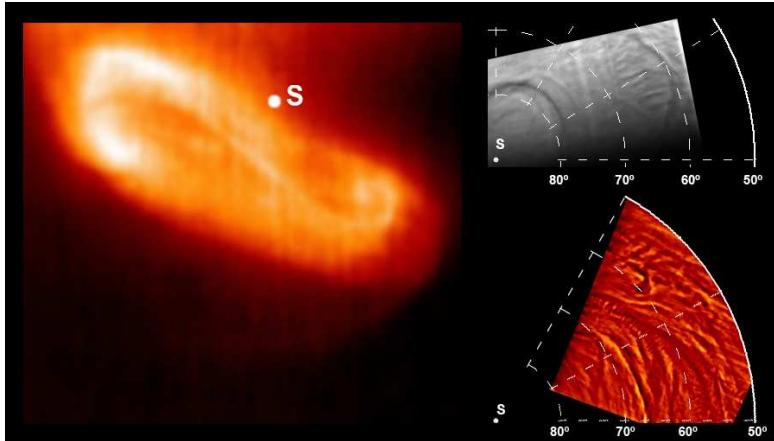


Figure 8: VIRTIS observations of the South polar dipole as seen at  $3.8 \mu\text{m}$  in the night-side (left, with the South pole location indicated by the S) and systems of gravity waves observed in the visible ( $380 \text{ nm}$ , right top) and the infrared ( $1.74 \mu\text{m}$ , right bottom). The  $380 \text{ nm}$  image (right top) also shows one side of the dipole on the upper cloud. Left and right figures not at scale

#### 4.4 Mesospheric dynamics

Above the clouds the atmospheric dynamics can also be studied by Venus Express by following the evolution of airglow emissions caused by a complex photochemistry at low atmospheric densities [27]. Among them, the  $\text{O}_2$  nightglow emission at  $1.27 \mu\text{m}$  is the most intense [28,29]. VIRTIS observations spatially resolve the airglow activity which is spectacularly variable not only in its morphology and intensity but also in the apparent motions of the airglow structures [30,31]. Visual tracking of the bright features allowed to obtain mean zonal and meridional motions related to a subsolar to antisolar circulation expected at mesospheric levels at an altitude range of 95-107 km [31]. Figure 9 shows a map of airglow activity as well as the inferred motions in the oxygen airglow structures. The zonal velocity is dominated by an intense prograde jet (contrary to the retrograde planetary and lower atmosphere rotation). Typical zonal velocities range between +60 (prograde) to -50 (retrograde) m/s, whereas most meridional velocities range from -20 (polewards) to +100 m/s (equatorwards) with an average meridional circulation of +20 m/s towards low latitudes.

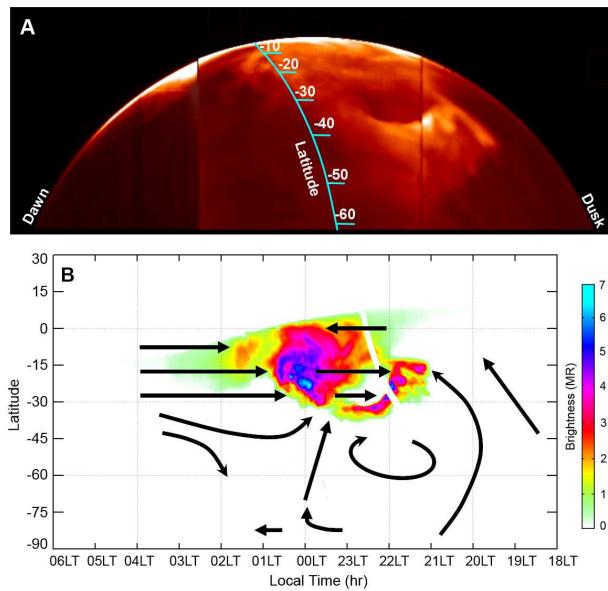


Figure 9: Airglow structures (bright patches) and average derived motions averaged over several Venus Express orbits. The airglow intensity largely increases at equatorial latitudes and specially at the antisolar point. Extracted from [27].

## 5 Conclusions

Venus Express is providing new insights of the atmosphere of this terrestrial planet. As the mission proceeds, it will produce a detailed view of atmospheric processes and dynamics over two full Venus years. Hopefully, the mission will be operative until the arrival of the Japanese Venus Climate Orbiter (also known as Planet-C) in 2009 that will enter an equatorial orbit of the planet. Longer term plans to continue the interrupted exploration of Venus are underway by ESA and NASA.

*Acknowledgements:* Venus Express is a mission of the European Space Agency. This work has been funded by Spanish MEC AYA2006-07735 with FEDER support and Grupos UPV 15946/2004 and Grupos GV IT464-07. RH acknowledges a “Ramón y Cajal” contract from MEC and JP a fellowship from UPV.

## References

- [1] Svedhem, H. et al. 2007, *Planet. Space Sci.* 55, 1636
- [2] Titov, D. V. et al. 2006, *Planet. Space Sci.* 54, 1279
- [3] Coradini, A. et al. 1999, *Adv. Space Res.* 24, 1105
- [4] Piccioni, G. et al. 2006, *ESA-SP*, 1291
- [5] Drossart, P. et al. 2007, *Planet. Space Sci.* 55, 1653
- [6] Esposito, L. W. et al. 1997, in “VENUS II: Geology, Geophysics, Atmosphere, and Solar Wind Environment”, University of Arizona Press
- [7] Esposito, L. W., et al. 1983, in “VENUS”, University of Arizona Press
- [8] Ragent B. et al. 1985. Particulate matter in the Venus atmosphere. *Adv. Space Res.* 5, 85
- [9] Hansen, J. E. and Travis, L. D. 1974, *Space Sci. Rev.* 16, 527
- [10] Ignatiev, N. I. et al. 1997, *Space Sci. Rev.* 45, 427
- [11] Belton et al. 1991 *Science* 253, 1531
- [12] Carlson, R. W., et al. 1991, *Science* 253, 1541
- [13] Kerzhanovich, V. V. and Limaye, S. S. 1985, *Adv. Space Res.* 5, 59
- [14] Preston R. A., et al. 1986, *Science* 231, 1414
- [15] Schubert. G. 1983, in “VENUS”, University of Arizona Press, 681
- [16] Gierasch, P. J. et al. 1997, in “VENUS II: Geology, Geophysics, Atmosphere, and Solar Wind Environment”, University of Arizona Press, 459
- [17] Limaye, S. S. and Suomi, V. E. 1981, *J. Atmos. Sci.* 38, 1220
- [18] Peralta, J., Hueso, R. and Sánchez-Lavega, A. 2007, *Icarus* 190, 469
- [19] Markiewicz, W. J. et al. 2007, *Nature* 450, 633
- [20] Sánchez-Lavega, A. et al. 2008, submitted to *Geophys. Res. Lett*
- [21] Del Genio, A. and W. B. Rossow 1990, *J. Atmos. Sci.* 47, 293
- [22] Smith, M. D., Gierasch, P. J., Schinder, P. J. 1993, *J. Atmos. Sci.* 50, 4080
- [23] Piccioni, G. et al. 2007, *Nature* 450, 637
- [24] Taylor, F. W. et al. 1979, *Science* 205, 65
- [25] Peralta, J., et al. 2008, *Procs. of the European Geosciences Union General Assembly 2008*
- [26] Russell, C. T. et al. 2007, *Nature* 450, 661
- [27] Bougher, S. W., Rafkin, S. and Drossart, P. 2006, *Planet. Space Sci.* 54, 1371
- [28] Crisp, D., et al. 1996, *J. Geophys. Res.* 101, 4577
- [29] Lellouch, E. 1997, in “Venus II: Geology, Geophysics, Atmospheres, and Solar Wind Environment”, The University of Arizona Press
- [30] Drossart, P. et al. 2007, *Nature* 450, 641
- [31] R. Hueso, et al. 2008, submitted to *Journal of Geophys. Res.: Planets*

# THE UPPER ATMOSPHERE OF VENUS OBSERVED BY VENUS EXPRESS

MIGUEL A. LÓPEZ-VALVERDE , GABRIELLA GILLI, MAYA GARCÍA-COMAS

*Departamento Sistema Solar, Instituto de Astrofísica de Andalucía–Consejo Superior de Investigaciones Científicas (CSIC), E-18008 Granada, SPAIN*

FRANCISCO GONZÁLEZ-GALINDO

*Laboratoire de Meteorologie Dynamique, Institut Pierre Simon Laplace, Paris, France*

PIERRE DROSSART

*LESIA, Observatoire de Paris-Meudon, UMR 8109 F-92190 Paris, France*

GIUSEPPE PICCIONI

*IASF-INAF, via del Fosso del Cavaliere 100, 00103 Rome, Italy*

AND THE VIRTIS/VENUS EXPRESS TEAM

**Abstract:** The upper atmosphere of Venus is still nowadays a highly unknown region in the scientific context of the terrestrial planetary atmospheres. The Earth's stratosphere and mesosphere continue being studied with increasingly sophisticated sounders and in-situ instrumentation [1, 2]. Also on Mars, its intensive on-going exploration is gathering a whole new set of data on its upper atmosphere [3, 4, 5]. On Venus, however, the only recent progress came from theoretical model developments, from ground observations and from revisits of past missions' data, like Pioneer Venus. More and new data are needed [6, 7, 8]. The arrival of the European Venus Express (VEX) mission at Venus on April 2006 marked the start of an exciting period with new data from a systematic sounding of the Venus atmosphere from orbit [9, 10]. A suite of diverse instrumentation is obtaining new observations of the atmosphere of Venus. After one and a half years in orbit, and although the data are still under validation and extensive analysis, first results are starting to be published. In addition to those global descriptions of VEX and its first achievements, we present here a review on what VEX data are adding to the exploration of this upper region of the atmosphere of Venus. We present measurements at those altitudes from one of the infrared sounders aboard VEX, the instrument VIRTIS, as an example of unique insights on the upper mesosphere and lower thermosphere of Venus, and discuss briefly the synergy with other instruments on VEX. We will conclude with our opinion on the importance and limitations of the Venus Express mission in order to broaden our global understanding of the upper layers of the terrestrial atmospheres.

**Keywords:** Venus Express – Planetary atmospheres – Remote sounding – Atmospheric dynamics – Airglow – Ionospheres – Non-LTE emissions

## 1 Introduction

A new air full of new data is flowing over the research on the atmospheres of the terrestrial planets, Mars and Venus. The second, in particular, has been severely affected by the lack of detailed observations from space for more than two decades. This is specially severe in the case of the upper atmosphere, the layers above the tropopause or equivalent layer (see Figure 1). The recent Venus Express mission is trying to change the trend, and its recent results promise to partially fill this gap.

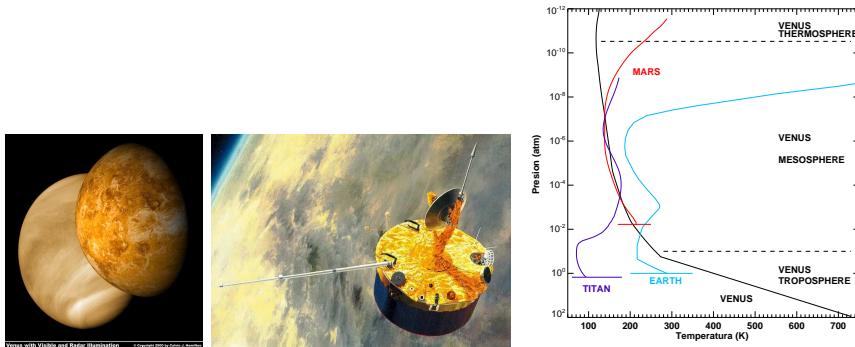


Figure 1: Left: Visible and radar views of Venus, from Mariner 10 and Magellan, mapping the cloud layer and the surface. Center: artistic vision of Pioneer Venus in orbit; credit: NASA. Right: Typical profiles of atmospheric temperature in Earth, Mars, Venus and Titan

About eight years ago, in a review on Comparative Aeronomy in the Solar System [11], a number of developments in atmospheric modeling of the upper atmospheres of the terrestrial planets were reported, but mostly based on the data from mission to Mars and Venus in the 70s and 80s. As an example, in the case of these planets' ionospheres, Nagy and Cravens mentioned in that review that the few we know about them comes from the Venus case, and from one single mission, Pioneer Venus (PVO, hereonafter), and much of what was found at that time was extrapolated to Mars given the expected similarity [12]. Only a few years later, however, a peculiar third ionospheric layer was found by the radio science experiment on Mars Express ([13]), and also that in contrast to Venus, a sudden ionopause seems to be absent, a result recently confirmed by Venus Express [14]. Also Bougher and colleagues concluded then [23] that in spite of the fruitful modeling of the basic parameters of the upper atmospheres of the terrestrial planets, like the governing patterns of their general circulation or their exospheric temperatures, the current models suffer from a "geocentric perspective" which should be revised once global measurements of com-

position, temperature and wind measurements are taken. These suffice to stress the need of data, ideally on a systematic basis, as a basic requirement for a sounded advance in our understanding of these layers.

However, sounding of the upper atmosphere is a difficult goal for several reasons. In-situ measurements, either from balloons or during the periapsis of satellite orbits, can normally study either the lower or the upper boundaries of the upper atmosphere. Most part of what we call the upper atmosphere, requires remote sounding techniques for its observation. Moreover the low densities at these altitudes demand a limb sounding. The lack of orbiters with technological capabilities for limb sounding has been for a long time a serious limitation. This technique was first applied to the investigation of the Earth's upper atmosphere in the late 70s with the NASA Nimbus series of satellites. Instruments on Nimbus 7 scrutinized our stratosphere for temperature and composition with a pointing and a precise inversion which were novelties at that time. These first limb radiometers were followed by a subsequent series of improved instrumentation on the Upper Atmosphere Research Satellite, in the 80s, with technological improvements but still mostly radiometers [16]. Another step in limb sounding of the Earth's atmosphere was performed in the 80s, with the use of high spectral resolution interferometers, like ATMOS, on the SpaceLab [17]. And in the 90s, with lighter versions which sounded even higher, like the MIPAS interferometer on board the European ENVISAT satellite [1, 18]. The wealth of data acquired by these missions include detailed photochemistry of the ozone hole [19], studies of the global dynamics of the stratosphere and the polar vortex [20], the detection of the atmospheric response to solar storms and energetic particles [21], the development of local climatologies, or the tracking of trends and climatic changes affecting the higher atmosphere [22, 18].

Also on Mars, the wave of missions to the red planet during the last decade has improved the study of the upper atmosphere significantly. This started in 1997 with the descent of Pathfinder through the Martian atmosphere, and continued with Mars Global Surveyor and Mars Odyssey orbiters, with exciting results at high altitudes. Details can be found in a companion paper in this issue [23]. There are also interesting results about the upper atmosphere of Mars obtained by Mars Express [5], a mission which will continue in operation until 2009, at least. This mission included some of the observational improvements which were so successful on the Earth's upper atmosphere, like limb sounding and a relatively high spectral resolution. This is the case of the Planetary Fourier Spectrometer, PFS [24], with limb capabilities in the infrared at a resolution of 1.5 cm<sup>-1</sup>, or the case of SPICAM [25], an stellar occultation experiment, which is sounding densities well above 140 km, or the case of the OMEGA instrument [26], with a fantastic vertical resolution at the limb (about 400 m [27]). The teams of these instruments are reporting the detection of auroral emissions on Mars [28], high altitude CO<sub>2</sub> ice clouds [29], or daytime fluorescence by CO<sub>2</sub> and CO [30]; the last ones occur at mesospheric and thermospheric layers and confirm at last

old predictions by non-local thermodynamic equilibrium (non-LTE) models [31, 32].

Research on the Venus atmosphere was certainly behind the developments in their neighbor planets. Demands of new data by the scientific community in the 90s [7, 33], together with interesting and new ground based observations [34, 36], and with the availability of optimal instrumentation from other missions (Rosetta and Mars Express), paved the way for ESA to take the initiative of promoting a mission to study Venus in the 2001 [37]. VEX was finally launched in November 2005, and represents, in the case of the study of the Venus atmosphere, the first mission designed for such purpose in more than 25 years.

This paper focuses on a review of the VEX mission and its ability to study the upper atmosphere of Venus. The term upper atmosphere is used freely by different scientific communities. In our work, this region is defined as that fraction of the Venus atmosphere laying above the cloud tops, above about 70 km. VEX is in normal scientific operations around our neighbor planet since June 2006. First findings have been recently published in a special issue in Nature [9]. This is an exciting time where new results are starting to emerge and to be published. We aim here at describing in a global sense how VEX instruments and observational strategy are prepared to contribute to our understanding of this special region in Venus, its upper atmosphere, adding to what is known so far about it. We start in section 2 with a small review of the current description of the Venus atmosphere, with basic results from previous observations and space missions. In section 3 we focus on its upper regions, scientific interest and particular problems which remain open. Then we describe briefly the VEX instrumentation, in section 4. We present results from one of these experiments, VIRTIS, in section 5, and in Section 6 discuss the synergy with other instruments on VEX. We will summarize our perspectives for the coming years of VEX in section 7.

## 2 Peculiarities of the atmosphere of Venus

The basic characteristics of the atmosphere of Venus can be reviewed from an historical perspective, following a selection of highlights of its observations.

Venus is sometimes considered as the twin planet of Earth, sharing similar size, density, formation age and distance from the Sun [10, 8]. However, strong differences exist between the two planets in atmospheric basic properties, these including very high temperatures and pressures at the surface, about 730 K and 90 bar, an almost pure CO<sub>2</sub> composition, and a sky fully and continuously covered by thick clouds with a complex layered structure containing sulfuric acid droplets. Also the planet's rotation is retrograde (177 degree inclination, hence no seasonal changes) and very slow (243 Earth's days), which nowadays is recognized as one key ingredient of the very different and peculiar atmospheric dynamics on Venus [39, 40]. These harsh conditions at the surface explain the fact that Venus research was neglected for a



Figure 2: Launch of VEX from Baikonur, November 25, 2005 (left panel), composition of pictures to depict its separation from the Fregat vehicle (central panel), and diagram of the limb sounding of the atmosphere by stellar occultation (right panel).

long time, given the impossibility to harbor eventual human exploration [33].

## 2.1 Microwave observations

In the 50s, during the early days of radioastronomy, strong microwave emissions indicated very high surface temperatures [38]. The rotation period was also detected by radio observations from Earth in the early 60s. This gave Venus the honorable first place in the objectives of space exploration of that time: the Mariner 2 mission was sent to Venus in 1962 [41]. Still nowadays radio ground-based observations of Venus are performed frequently, which are useful to sound the Venus mesosphere. These observations are supplying winds and distribution of minor compounds like CO at those altitudes [42]. During the Venus Express mission microwave observations are particularly useful for their addition of complementary observations, and were part of a recent VEX validation campaign by ground-based observations [45].

After the first microwave measurements confirmed the rotation of the solid body, the superrotation of the Venus atmosphere was established. Still the origin and mechanisms of this effect is an open issue, with debatable and alternative theories. The general view is that it is induced or related to the slow rotation of the planet, combined with some effective transfer of momentum from the lower atmosphere to the clouds' altitude.

## 2.2 Spacecrafts and landers

More than 30 missions have made the trip to Venus [9], including the Mariner series of fly-by from 1962 to 1975, the Venera series (descent probes with landers) from 1967 to 1985, the Pioneer Venus mission (one orbiter and four probes) from 1978 to 1982, the two Vega missions (including 2 balloon stations and 2 landers) on 1985, the

American orbiter Magellan from 1990 to 1994, and the fly-by by Galileo and Cassini in 1991 and 2001, respectively. Reviews of all of them can be found elsewhere [8].

The atmospheric composition, density and thermal structure are essential properties of an atmosphere, and were among the first objectives of the first missions. The first probes plunged into the Venus atmosphere found almost identical thermal profiles, with few horizontal variations [44]. This was not surprising, given the large density of the atmosphere; the mass of the Venus atmosphere should store and distribute heat efficiently. In those early years of Venus exploration there were serious problems to understand the greenhouse effect observed, the high temperatures at the surface. After the first determinations of the precise composition of the atmosphere, it was clear that such a CO<sub>2</sub>-dominated atmosphere should have numerous gaps in the infrared where radiation could escape to space, and consequently, cool the atmosphere. The small amounts of water vapor present below the clouds did not close those infrared windows [7]. A quantitative understanding only started much later, after the composition of the top clouds was determined by the series of probes (H<sub>2</sub>SO<sub>4</sub> droplets are efficient absorbers in the near-IR and highly transparent in the visible), after advances in spectroscopic high temperature databases, and after an improved treatment of line shape effects into radiative transfer models [47].

A lot of what we know about the Venus atmosphere comes from the Pioneer Venus mission, orbiter and probes, which reached Venus in 1978. One of the many innovative investigations from the orbiter was to use cameras at different wavelengths for the study of the clouds and the surface emissivity. These confirmed the superrotation of the lower atmosphere and found that the maximum occurred at the altitude of the clouds. Their measurement of radiative fluxes at the cloud deck demonstrated that half the solar radiation absorbed in the Venus atmosphere is actually deposited there, by some efficient and unknown absorber in the UV [7]. Many questions still remain regarding the clouds, like their actual role in driving the circulation at those altitudes. They may merely respond to microphysical processes and photochemical reactions, which might be highly independent on the atmosphere's movements.

A number of key measurements of the atmosphere of Venus above the clouds were obtained by PVO, including the thermal structure, the dynamics of the mesosphere, and the ionosphere. We discuss them in the next section.

### 3 The upper atmosphere of Venus.

The upper layers of an atmosphere are exciting from various viewpoints or disciplines in planetary science. Diverse escape processes occur there and constitute one key factor in the long term evolution of the atmosphere and of the planet. The interaction with the solar wind is strongest at these altitudes; the solar radiation is filtered and partially transmitted to lower regions, and therefore, the conditions below and at the

planet's surface depend on the densities and compositions higher up. In addition, these peculiar conditions offer an excellent benchmark to study specific chemical and physical processes which test our understanding in fields like spectroscopy, radiative transfer, convection and hydrodynamics, to mention a few. These layers can supply also useful information to infer geophysical properties of the planet, like its internal structure, and to supply correct boundary conditions for the general dynamics and chemistry of the whole atmosphere. An advantage of the outermost layers is their accessibility, sometimes better than lower layers; they can be reached by low orbiting satellites, and by the remote sounding of the different airglow emissions produced at these altitudes. Describing in detail those emissions and using these layer's densities to modify the satellites orbits are also among the objectives of the research on the upper atmosphere. All these conditions concur in the case of the Venus thermosphere, and have been investigated with the data available so far. In this section we review briefly a selection of the results and problems which require further measurements.

### 3.1 Thermal structure

The thermal structure obtained by radio occultation and by the PVO infrared radiometer and probes, responded to convective and radiative processes up to about 90 km, as expected [44, 46, 47], but showed peculiar latitudinal variations, with a positive gradient towards the poles at altitude between 70 and 90 km [48]. At high latitudes the atmosphere at that layer is well stratified, with a colder region below, termed the cold collar, and which seems to surround the pole [52]. It is not known if this cold collar is frequent or not (Magellan radio occultation did not reveal it [60]).

A transition to the upper atmosphere, with an isothermal or inversion layer equivalent to the Earth's tropopause, above which the energy balance is controlled by radiation, is not as well marked in Venus as it is on Earth. Normally the Venus mesosphere is simply taken as that region from the cloud top, to the base of the thermosphere, typically at 0.1  $\mu$ b, around or above 120 km. One-dimensional global models predicted a mesopeak, produced by solar absorption in CO<sub>2</sub> near-IR bands and located around 1  $\mu$ b, or about 100 km altitude. This layer seems to be confirmed by ground-based microwave observations, and has been used as a different definition of the mesopause, [43], since large variability is observed above it. But it has not been described in a systematic manner from instrumentation in orbit. An interesting result reported at those altitudes from SPICAM on VEX, and not observed previously, is a strong inversion layer reported there [51], which opens an additional debate about its source and its relation with the mesopeak.

The cold collar mentioned above and tidal waves observed on Venus upper atmosphere, are deviations from the radiative equilibrium situation [44], and which might be partly related to the superrotation; they present challenges to future global modelling of the Venus atmosphere. Also, it is not known why large latitudinal variations

are present while the longitudinal changes, following similar solar illumination, are very modest [53]. Among the opened questions, from the observational point of view, are the confirmation of these structures, and the study of their repeatability and variability in both hemispheres.

In the thermosphere, PVO observations show relatively mild dayside temperatures, and very cold nighttime temperatures, what is referred to as the Venus cryosphere. This was a puzzle for a long time, in view of the proximity to the Sun. A solution, at least partial, was found in the early 90s. Reanalysis of non-LTE CO<sub>2</sub> measurements in the Earth's upper atmosphere [55, 56] revealed a much more powerful role of CO<sub>2</sub> as a cooling agent of the atmosphere, due to a more efficient energy transfer between CO<sub>2</sub> and atomic oxygen than previously assumed. This explained the strong buffering of solar variations by the Venus thermosphere, explaining the low response to the solar cycle. Still there are uncertainties in this rate coefficient, specially at low temperatures, where measurements in laboratory are difficult [58].

### 3.2 Dynamics, waves and large scale transport

Some reviews examined the fundamental problems in the atmospheric dynamics of Venus below the clouds and in the lower mesosphere [39]. A companion paper in this issue also tackles this topic [35]. A number of works summarize also our current understanding of the dynamics of the upper atmosphere [58]. We present here our own view of the most interesting problems at these altitudes.

Above the cloud top the circulation has been linked to the thermal structure observed, using the assumption of cyclostrophic balance [54, 57]. Nevertheless, a systematic sounding of temperatures would be needed to establish the extent of this approximation. The situation was originally described by one Hadley cell type of circulation, raising at mid latitudes and cooling the mesosphere, then moving towards the poles in the upper mesosphere, and descending at the poles; the returning would take place at altitudes below the clouds [52]. Actually, a revision with better radiative equilibrium models one decade later revealed key details of the nature of this mesospheric circulation [39, 46]. As this circulation would be against the radiative balance, it must be thermodynamically induced in an indirect manner, perhaps from an interaction of the tides with the mean flow of the zonal superrotation, this being retarded by them [47]. Indeed, small longitudinal variations were detected by PVO in the upper mesosphere, from 5 to 10 K, which indicate tidal propagation [39]. How this is combined with the cyclostrophic balance, and how it merges with the thermospheric circulation is not well known. The altitude decrease of the superrotation might be simply due to the increasing solar warming at higher altitudes [10]. Few data from microwave ground-based observations suggest that, at the 100 km level, 1-week long changes occur between both regimes [43]. In order to clarify this topic, systematic sounding of temperatures and winds, in the mesosphere, together with

precise radiative models, would be very beneficial.

At thermospheric altitudes, open questions also exist. The strong temperature gradients between dayside and nightside seem to drive an interhemispheric large scale circulation. The winds, however, seem to be lower than the expected 400 m/s value, which indicates that some retarding mechanism must be operating [58]. It is expected that the interhemispheric flow and the rising and downwelling branches of such circulation will affect the transport of minor species and might produce some diabatic heating. Such heating has very recently conjectured as responsible for inversion layers in the nightside observed by SPICAM on VEX [51]. Evidences were also found in NO airglow emissions, showing an increase in the nightside thermosphere, approximately where the subsidence may occur [58]. It has also been suggested that this downwelling may enrich the nighttime mesosphere in atomic oxygen, in order to explain the strong emission by molecular oxygen observed by ground-based telescopes. These observations, at 1.27  $\mu\text{m}$ , showed large spatial and temporal variations, in addition to very high levels of emission [34]. The usual explanation, under debate in its details, include downwelling of atomic oxygen followed by recombination. Convection, or some transient dynamical phenomena, like wave breaking might play a role in its variability [49]. New data from VEX is starting to clarify the situation [59].

The existence of wave activity at high altitudes of the Venus atmosphere has been speculated for a long time, but not much observational evidence was available. One set of data showing clear signs of gravity waves come from the few radio occultation by Magellan [60]. Another evidence reported [61] is the perturbation in densities of  $\text{CO}_2$ , O and He detected by the neutral mass spectrometer on PVO, which show larger amplitudes during nighttime than during daytime, and this is supported by a model of gravity wave propagation [62]. Another interesting example comes also from the Magellan mission, but from a very recent analysis of the radar tracking of the orbit [63], showing for the first time a 9-day period oscillation in the thermosphere, and which seems to be restricted between dusk and midnight. The lack of previous detection suggests, in these authors' opinion, either peculiar modes of propagation and interaction with the mean flow, or perhaps the existence of excitation sources above the cloud tops. The wavy nature of the upper atmosphere is a very open topic, and is hoped that missions like Venus Express may shed some light on it [63].

Transport of minor gases is a useful mean to track the dynamics of the atmosphere. In the lower atmosphere, latitudinal gradients of CO have been observed [64]. An explanation makes use of the circulation in the lower atmosphere, but the strong mixing expected there might preclude the gradient. The alternative is that descent from the CO-richer upper atmosphere may occur in the polar regions, as it happens on the Earth middle and upper atmosphere [65, 7]. Transport processes like these are, however, speculative and remain to be confirmed by new measurements.

### 3.3 Non-LTE emissions

Airglow and non-LTE emissions are among the characteristic features of the upper atmosphere. They usually supply direct diagnostic tools of the peculiar chemistry and physics at those altitudes, permit remote sounding of atmospheric temperatures and abundances, and via numerical models, drive the investigation of the energy balance of the atmosphere.

While a number of emissions are worth examining, we focus here on two specially interesting ones, for its potential for sounding the Venus upper atmosphere and because new light will be shed upon them after VEX.

Full understanding of these emissions requires careful 1-D modelling, including radiative transfer, microscopic physics, and spectroscopic databases. Sometimes, optically thin conditions facilitate their analysis. This is not the case of the intense non-LTE emissions by the major species of the Venus atmosphere, CO<sub>2</sub>, as first pointed out by the pioneering radiative transfer modelling for Venus by Dickinson, in 1972 [66]. He realized the important role played by minor bands of CO<sub>2</sub> in the radiative balance of the mesosphere, and was the base of later model developments [67, 68, 69]. These came after the detection of strong infrared emissions from Venus and Mars, in the CO<sub>2</sub> laser bands at 10  $\mu\text{m}$  [70, 71], and which were explained by solar fluorescence of the fundamental band of CO<sub>2</sub> at 4.3  $\mu\text{m}$ , followed by radiative relaxation at 10  $\mu\text{m}$ . During the last decade, and motivated by the prospect of IR sounding of the Martian atmosphere [72], more comprehensive non-LTE models were developed, including a larger number of CO<sub>2</sub> emissions covering from 1 to 20  $\mu\text{m}$  [31, 32], which were later extended to the Venus atmosphere [73]. These efforts supply tools to analyze remote measurements, and quantify the radiative balance of the Venus atmosphere, dominated by the CO<sub>2</sub> infrared emissions in the upper mesosphere and lower thermosphere. However, few data so far were available for the validation of these models. A few limb spectra at 4.3  $\mu\text{m}$  were taken during the Galileo fly-by with the NIMS instrument, and explained recently, in preparation for the VEX data analysis [74]. Fortunately, the instrument VIRTIS is detecting non-LTE CO<sub>2</sub> radiances systematically [59], which are helping to test the model and will hopefully be used to study the upper atmosphere [75].

Another important emission is the molecular oxygen's infrared system, at 1.27  $\mu\text{m}$ , mentioned above. Its photochemistry is well known, but the above mentioned measurements puzzled the theory for a long time, unable to explain the peak emission and its strong variability. New data from the instrument VIRTIS/VEX have much better spatial coverage than ground-based measurements, and they are starting to clarify some of the issues, like the strong signal which seems to fit the theory [59]. The systematic sounding with VIRTIS is currently characterizing the morphology and variability of the O<sub>2</sub> emitting regions [76, 35].

### 3.4 Ionospheric layers and atmospheric escape

The Venus atmosphere and surface are very dry and, assuming initial inventories similar to present Earth, an important loss of water, in the form of hydrogen escape and more important than on Earth, must have taken place on Venus during its evolution. This is supported by the about 100 times higher D/H ratio measured in Venus [77], showing a preferential loss of the heavier isotope via thermal escape. The loss rate estimated for Venus suggests that an Earth ocean could have been lost in about 100 million years [79]. The current escape rate is thought to be much smaller in Venus now, and must surely be by non-thermal mechanisms.

The Venus ionosphere was well described after radio occultation and in-situ data by PVO, and its basic characteristics, like an ionopause between 300 and 400 km varying with the solar activity, the nature of the primary ions ( $O_2$ ,  $O$  and  $CO_2$ ) and the altitude (around 140 km) and density of the electron peak [50]. Most of these results were well reproduced by models [50, 12] although many unknowns remained, like the role played by the solar wind in accelerating ions which could contribute to an important atmospheric escape. PVO measurements correspond to solar maximum conditions, and what the ionosphere looks like at solar minimum was unknown until Venus Express. VEX's magnetometer (MAG) is carrying out regular in-situ measurements of the near-Venus plasma during the orbital periapsis (about 250-350 km). It also has much better plasma analyzers than PVO. The first results show that the bow shock reflects a fully magnetized ionosphere, with no penetration of the solar wind into the ionosphere [80]. This may mean that no strong solar wind interaction is present [10], although according to the magnetometer team, the magnetic field observed might accelerate the ionospheric particles in the night tail, inducing atmospheric losses there. They suggest to confirm this by measurements of plasma velocities. Loss of ions has been observed by VEX by the plasma analyzer, at the boundary of the induced magnetosphere. These include  $O^+$  and  $H^+$ , primarily, and their relative loss rates are close to the 2/1 value which would indicate water photolysis as the main source. This seems to indicate that there is an escaping mechanism operating at present in Venus, acting on the ions rather than on the neutrals [81].

## 4 The Venus Express mission.

In the previous section we have illustrated some key problems in our understanding of the known properties observed in the Venus atmosphere. VEX is equipped by a series of instruments with innovative characteristics which make them ideal for sounding the upper atmosphere of Venus, and which offer excellent chances to study some of the open problems listed above. Summaries of all the different instruments, goals and characteristics are given in detail elsewhere [45].

Let us recall that the driving idea of the whole VEX mission, and its design and

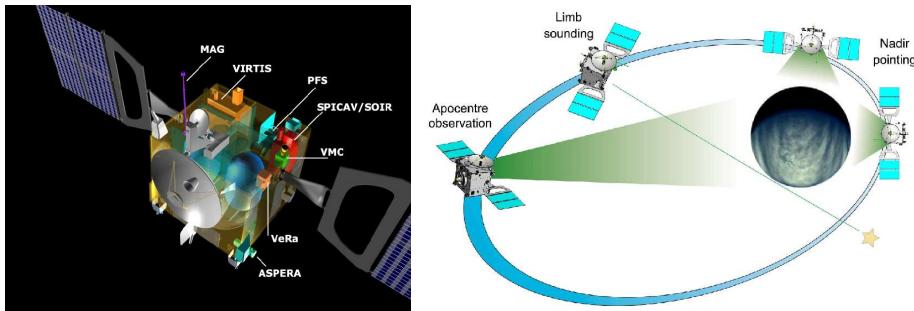


Figure 3: Diagrams showing the accommodation of the scientific payload of VEX on the spacecraft, and the observational phases. Credit: ESA

implementation at ESA have been very special, with a time record between concept (2001) and launch (2005). This was thanks to the use of a suite of instruments taken from previous missions, MEX and Rosetta, quickly assembled into a Mars Express platform, and launched with a Soyuz-Fregat to reach a very elliptical orbit on Venus, similar to that of Mars Express around Mars. A realistic road map of observations, with precise scheduling of the operational time for each instrument during the VEX orbit, is available elsewhere [45].

Figure 4 lists the VEX instruments, with their modes of operation and main scientific objectives. Figure 3 shows the accommodation of the instruments on the spacecraft. Some words about the VEX orbit are necessary. Each orbit lasts 24 hours, with about 10 hours devoted to data transmission to Earth. Some instruments are specially appropriate to operate near periaresis and others can observe from farther away. The actual Venus measurement period can be divided in two modes of observation, the imaging mode and the spectral mode, when the spacecraft is closer/farther than 12000 km from the planet, respectively. This is a relevant division for those instruments like VIRTIS, which can adapt their observation for example from nadir global images at low spectral resolution to limb sounding with higher spectral resolution.

VEX has a number of clear advantages to previous Venus missions. First of all, one is the technological improvement in all the instruments, compared to similar experiments about 20 years ago, including also the larger data transfer rates. Second, the better spacecraft 3-axis stability, permitting precise pointing, essential for limb sounding, and imaging spectroscopy; these were impossible with a spinning satellite like PVO. Third, new techniques are applied for the first time in orbit around Venus, like the “energetic neutral atom” imaging [37], or solar and stellar occultations. Four, the near polar orbit permits sounding of regions hardly accessible by previous mis-

sions, or from Earth observations. The large ellipticity of the orbit, which impose stringent limitations for close mapping and limb sounding, is an advantage however for a number of studies. These include image tracking and global viewing, which may permit observations of planetary waves, polar vortex studies, and time evolution of dynamical features, like winds and turbulence; as well as low periapsis passes, for in-situ exploration of the upper thermosphere.

Acronym	Original	Description	Scientific Objectives	Observation mode	Comment & advantages
ASPERA	MEX	Analyzer of Space Plasmas and Energetic Atoms	Neutrals (H,O), ions and electrons. Discriminate masses and angul.distrib	Entire orbit. Nadir at close approach	New technique in Venus
SPICAM /SOIR	MEX	UV/IR spectrometer for solar/stellar occultation	UV: densities upper atm IR: H <sub>2</sub> O,D/H,SO <sub>2</sub> ,hazes	Near Pericenter Limb	First time in Venus
PFS	MEX	High resolution IR Fourier Spectrometer	Full near-IR spectrum (1-5 $\mu$ m), temp,compos.	Near pericenter Nadir & Limb	Not operating
VERA	Rosetta	Radio science instrument	Ionosphere Mesosphere: density&temp Surface properties	Entire orbit Nadir & Limb	New technique in Venus
VIRTIS	Rosetta	UV-visible-IR Imaging Spectrometer	Composition Mesosphere: temp struct. Airglow & non-LTE	Entire orbit Nadir & Limb	New technique in Venus. Improves upon NIMS&VIMS
VMC	MEX	UV-visible Camera	Images/movies of clouds and waves; UV absorber	Entire orbit Nadir	MEX's was monochromatic Space&time resolution >PVO

Figure 4: List of VEX instruments, description and objectives

## 5 Exploring the upper atmosphere with VIRTIS

Here we present new data on the upper atmosphere of Venus by the instrument VIRTIS. We focus on limb observations, which are specially adequate for such purpose, and also on fluorescent emissions by CO<sub>2</sub> at 4.3  $\mu$ m, and their comparison with theoretical model predictions.

VIRTIS is operated to observe Venus in nadir/disc viewing when VEX is far from Venus, but also in a limb sounding geometry during periapsis. Although the large ellipticity of VEX takes VIRTIS as far as 60000 km away, the images from apoapsis can still be used to observe the atmospheric limb of the planet. Figure 5 shows a portion of the Venus disc, acquired during orbit #25, near the apoapsis of the VEX orbit. The

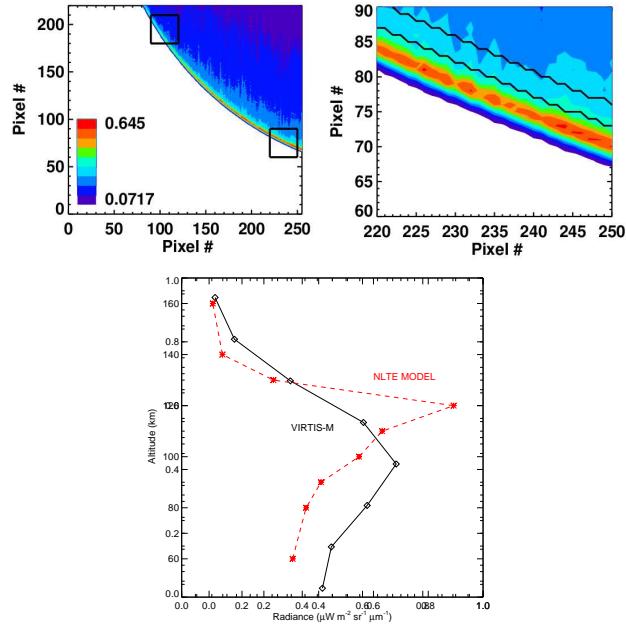


Figure 5: Left panel: Map of CO<sub>2</sub> radiances from VIRTIS-M, from Orbit#25, from apoapsis. Right panel: Zoom to detail the limb sounding capabilities. See text. Lower panel: Radiance profile and model simulation

radiance map corresponds to  $4.32 \mu\text{m}$ , and clearly shows a limb brightening, colored here in red. A zoom of the frame illustrates more clearly the limb emission. The pixel size is depicted by the oscillation in the two reference lines added, to mark the surface and the 60 km tangent altitude. This shows a clear peak emission around 120 km, and at a relatively coarse vertical resolution, but still, similar to the only other detection of this emission in Venus, from NIMS/Galileo. The horizontal resolution is much better, and represents a unique view of this emission. The peak intensity and altitude location, the vertical profile, and the solar illumination variations are a number of features which pose challenges to theoretical simulations.

We built a vertical profile at the limb from this V-M image, at one particular location and at one of the central wavelengths of the CO<sub>2</sub> emission band, at  $4.28 \mu\text{m}$ . It is shown in Figure 5. The model simulation, for a similar solar zenith angle and for a typical atmospheric temperature/density profile, is also shown in the figure. Any wavelength can be used, and [59] present a similar comparison at  $4.33 \mu\text{m}$ . The data seem to follow the model prediction, both the shape of the profile and the peak's

intensity and altitude. The responsible for this emission, the first excited state of  $\text{CO}_2$  in its asymmetric-stretching mode of vibration,  $\text{CO}_2(001)$ , reaches a very high solar pumping above about 120 km, according to the model. Below that altitude, the solar flux at  $4.3 \mu\text{m}$  is severely absorbed, and the emission decreases when pointing at the lower mesosphere, or below. At higher altitudes, the  $\text{CO}_2(001)$  relative population is about constant, but the total density decreases exponentially, and the  $\text{CO}_2$  emission therefore decreases quickly as well. A quantitative fit, however, requires a quantitative determination of the density structure of the Venus atmosphere, which is not known at this point. Alternatively, once the non-LTE model is validated, it could be used to retrieve the density, by fitting these emission profiles.

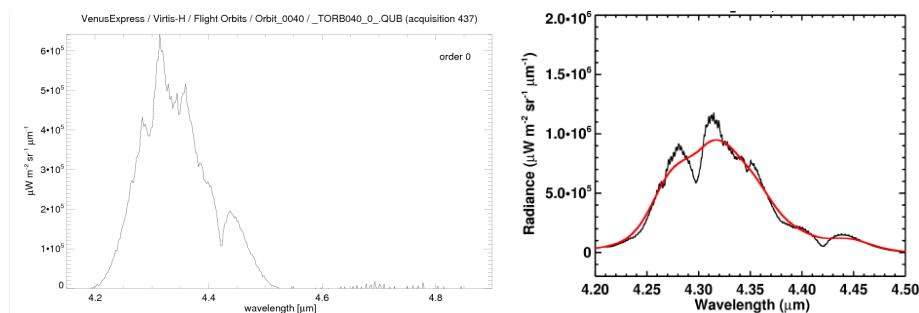


Figure 6: Left: Limb spectra from V-H at lower thermospheric altitudes. Right: Model simulation for V-H (black) and V-M (red). See text.

Figure 5 is an example from one of the signals of VIRTIS, at a resolving power  $R \sim 400$ , commonly used to produce global images or mosaics of Venus, and which is termed V-M. The other signal, at a higher resolution ( $R \sim 1800$ ) is termed V-H, and Figure 6 shows one particular spectrum taken in this mode, acquired during periapsis pointing at about 120 km. Several interesting features can be observed above the noise level, which is much lower, at about  $5000 \mu\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$  confirming expectations after ground calibration [37]. Non-LTE model simulations of VIRTIS spectra [74] are also shown in the figure, and the agreement is good too. As studied by [74], a large number of bands conform the limb emission, with weak and isotopic  $\text{CO}_2$  bands giving significant contributions in the wings of this spectral region. There seems to be, however, a portion of the spectrum which is not well represented by the model in absolute terms. The spectrum in the longward wavelength wing of the strong central emission, around  $4.40 \mu\text{m}$ , seems to be smaller in the model. This problem seems to be systematic, in all the V-H spectra studied so far. We attribute this to a number of weak bands, surely hot and combinational transitions which are not included in the non-LTE model. They may not be relevant to the energy budget of the upper

atmosphere, but a good fit of the measurements require to revise the non-LTE model by adding those minor transitions. Alternatively, some extra excitation source, still unidentified, may be present. This work is in progress.

One of our aims is to perform a detailed analysis of the emission and their different contributing bands. Such study may result in model improvements, like modifications in the way the model describes how the solar energy is transferred between the different CO<sub>2</sub> vibrational states, which includes collisional processes not well determined yet. There are hopes also that these high resolution data from VIRTIS may teach us about some microscopic energy transfer mechanisms.

Other interesting limb atmospheric results from VIRTIS regard the nighttime emissions of O<sub>2</sub> at 1.27 μm, which indicate time and spatial variations at high resolution for the first time. This is the topic of a companion work in this issue [35].

## 6 Synergy between Venus Express instruments

VIRTIS is not the only instrument on board VEX to study the upper atmosphere (see Figure 4). Some of them will supply useful information for direct and indirect correlation with VIRTIS data, and vice-versa.

A key progress by VEX is expected in the description of the global temperature and density structure of the upper atmosphere, including tracking of minor neutral species and ions. These will be tackled directly by SPICAV using stellar and solar occultation, by VERA using radio occultation, by ASPERA with in-situ data, and indirectly by VIRTIS, via the strong non-LTE emissions there; these are affected by the actual density and the distribution of the emitting species.

The VERA team plans to assume cyclostrophic balance to use the meridional distribution of atmospheric temperature to obtain the zonal wind distribution. The first results from this instrument show neat temperature profiles, with very small measurement noise, from below the cloud deck, around 45 km, up to about 90 km.

The upper boundary of VERA is about the lower range of temperature sounding for SPICAV, this extending up to at least 140 km altitude, as reported recently [51]. Although the results seem to match, there is a intriguing result from SPICAV, a systematic warm layer just above its boundary layer, around 100 km. Previous data from PVO and microwave observations also have upper boundaries about 100 km, with not good quality determinations. This altitude range is therefore subject to specially difficult measurements. Those other previous determinations suggest some small heating, but nothing like the more than 40 K obtained by SPICAV in some profiles. If SPICAV data are confirmed, some strong heating mechanism may be needed there. Since the data presented correspond to nightside, the SPICAV team suggest the heating may be produced by adiabatic compression in the downwelling branch of the global thermospheric circulation, something simulated previously by

models, but with not such intensity [23]. An argument in favor of this interpretation is the increasing magnitude of the inversion as the solar zenithal angle (SZA) increases (closer to the anti-solar point). They also extend their results to estimate a downward velocity for the total atmosphere; their estimation does not include, however, non-LTE effects and is therefore overestimated. However, the data contains the information for a more detail calculation.

Regarding the correlation with VIRTIS, present validation activities and data analysis are focused on the characterization of the emissions. The foreseeable derivation of densities from the CO<sub>2</sub> non-LTE emissions will still take some time. Strong emissions from VIRTIS are observed from the cloud tops upward, even up to above 160 km in some individual spectra. However, eventual retrievals will have to cope with opacity properties and will surely be useful only around the peak emission and upward. As the VERA and SPICAV densities are already available, we plan to use these as input to our model simulations for VIRTIS, which is an indirect comparison strategy between them.

Ideally, a direct correlation between all three instruments should be performed on individual profiles. However, this is difficult given the different observational requirements by individual instruments. VIRTIS, for example, is detecting solar pumping emissions, while most of SPICAV data correspond to nighttime stellar occultations. SOIR data are more interesting, but also occur at high SZA, nominally at 90 degree, where the CO<sub>2</sub> excitation is low. Therefore, each instrument will build its own global maps of densities, and these should be compared.

## 7 Summary and perspectives

A new mission to Venus, the European VEX, is at last in orbit around Venus, obtaining exciting new data and results for about one year and half to date. First analysis of the data are starting to be published, although the validation activity and systematic analysis of the data is still on-going.

At this early stage, overviews like the one presented here, focused on the atmospheric layers above the cloud tops, may (and hopefully, will) be obsolete soon, as new results are emerging quickly. Regardless of whichever specific results are to be found in the near future, we are listing and discussing briefly a number of scientific problems of the upper atmosphere of Venus which require new data, and we are framing some of our hopes on the new techniques and observational abilities of the VEX mission.

Regarding the upper atmosphere, three instruments, SPICAV, VERA, and VIRTIS are specially prone to correlative measurements of the density and temperature structure of this region. The emissions of VIRTIS, under study, are giving for the first time, a detailed picture with spatial and time variations, which confirm model predictions and offer chances to learn about density variations, turbulence and transport,

and even microscopic physics in the upper atmosphere. The density and temperature profiles from SPICAV and VERA are supplying new results with unprecedented detail of the vertical variation, the dynamics at global scale, and the extension of the cloud deck and its impact on the thermal budget. No doubt these new data from VEX in the mesosphere and thermosphere will trigger modeling efforts to understand them in the near future. Some of these efforts are already on-going, with the development of a number of global circulation models, and the offer to create a climate database similar to the successful one we are building on Mars nowadays [78].

Routine operations of VEX are granted by ESA until 2009. Extension of the VEX lifetime beyond 2009 is highly desirable. Not only Venus is under a different segment of the solar cycle, but a number of studies would be benefited from it. One is the correlation between different instruments, as mentioned above. A second one is the possibility to modify the VEX orbit, by using some fuel on board, which would supply new geographical perspectives of Venus. Most important, the scientific value of the new measurements merit such an extension. The present database promise already to keep us actively analyzing the data and testing our modelling tools for a long time. Also, the sister mission Mars Express is also operative on Mars at present, and with similar instrumentation, and this multiplies the scientific benefit from each of these missions. This is a unique opportunity for comparative planetology, and keeping both running would be a high priority for ESA and the European scientific community.

*Acknowledgments:* This work has been carried out under project ESP2004-01556.

## References

- [1] Laur, H. 2005, COSPAR Information Bulletin 162, 31
- [2] Mertens, C.J., Schmidlin, F.J., Goldberg, R.A. et al. 2004 GRL 31, L03105, doi:10.1029/2003GL018605
- [3] Albee, A.L., Palluconi, F.D., Arvidson, R.E. 1998, Science 279, 1671
- [4] Saunders, R.S., Arvidson, R.E. and Badhwar, G.D. 2004, Space Sci. Rev. 110, 1
- [5] Chicarro, A.F. 2006, Highlights Astron. 14, Proceedings of the IAU XXVI General Assembly 2006, CUP, 2007, 334
- [6] Crisp, D., Allen, M., Anicidh, V.G. et al. 2002, ASP Conference Proceedings, Vol. 272, "The Future of Solar System Exploration (2003-2013)", Astronomical Society of the Pacific, 5
- [7] Taylor, F.W. et al. 1999, "The Venusian Environment", ESA Technical Report
- [8] Titov, D.V., Svedhem, H. and Taylor, F.W. 2006, "The Atmosphere of Venus: Current Knowledge and Future Investigations" Springer Physics and Astronomy, London
- [9] Svedhem, H., Titov, D.V., Taylor, F.W. et al. 2007, Nature 450, 629
- [10] Ingersoll, A.P. 2007, Nature 450, 617
- [11] Mendillo, M., Nagy, A. and Waite, J.H. 2000, Editors, "Atmospheres in the solar system: Comparative aeronomy", Geophysical monograph. AGU, Washington

- [12] Nagy, A. and Cravens, T.E. 2000, in “Atmospheres in the solar system: Comparative aeronomy”, Eds: Mendillo, M., Nagy, A. and Waite, J.H. Geophysical monograph, AGU, Washington, 39
- [13] Pätzold, M., Tellmann, S., Hausler, B. et al. 2005, *Science* 310, 837
- [14] Pätzold, M., Häusler, B., Bird, M.K. et al. 2007, *Nature* 450, 657
- [15] Bougher, S.W., Roble, R.G. and Fuller-Rowell, T. 2000, in “Atmospheres in the solar system: Comparative aeronomy”, Eds: Mendillo, M., Nagy, A. and Waite, J.H., Geophysical monograph. AGU, Washington, 261
- [16] Reber, C.A. 1993, *GRL* 20, 1215
- [17] Gunson, M.R., Abbas, M., Abrams, M.C. et al. 1996, *GRL* 23, 2333
- [18] Fischer, H., Birk, M., Blom, C. et al. 2007, *Atmos. Chem. Phys. Discuss.* 7, 8795
- [19] Brachera, A., Bovensmanna, H., Bramstedta, K. et al. 2005, *ASR* 36, 855
- [20] López-Puertas, M., Funke, B., Clarmann, T. Von et al. 2006, *Space Sci. Rev.* 125, 403
- [21] Gardner, J.L., Lopez-Puertas, M., Funke, B. et al. 2005, *JGR* 110, doi:10.1029/2004JA010937
- [22] Thomas, G.E. 2003, *ASR* 32, 1737
- [23] González-Galindo, F., Forget, F., Angelats i Coll, M. et al. 2008, *LNEA* vol. 3, this issue, submitted
- [24] Formisano, V., Angrilli, F., Arnold, G. et al. 2005, *PSS* 53, 963
- [25] Bertaux, J.L., Korablev, O., Perrier, S. et al. 2006, *JGR* 111, doi:10.1029/2006JE002690
- [26] Bibring, J.P. 2007, *JGR* 112, doi:10.1029/2007JE002935
- [27] Drossart, P., López-Valverde, M.A., Comas-Garcia, M. et al. 2006, “2nd workshop on Mars atmosphere modelling and observations”, held February 27 - March 3, 2006 Granada, Spain.
- [28] Bertaux, J.L., Leblanc, F., Witasse, O. et al. 2005, *Nature* 435, 790
- [29] Montmessin, F., Bertaux, J.L., Quimerais, E. 2006, *Icarus* 183, 403
- [30] Formisano, V., Maturilli, A., Giuranna, M. et al. 2006, *Icarus* 182, 51
- [31] López-Valverde, M.A. 1990, *Emisiones infrarrojas en la atmosfera de Marte*, Ph D Thesis, University of Granada, Granada, Spain
- [32] López-Valverde, M.A. and López-Puertas, M. 1994, *JGR* 99, 13093
- [33] Hand, E. 2007, *Nature* 450, 607
- [34] Allen, D., Crisp, D., Meadows, V. 1992, *Nature* 359, 516
- [35] Hueso, R. et al. 2008, *LNEA* vol 3, this issue, submitted
- [36] Meadows, V.S., Crisp, D. 1996, *JGR* 101, 4595
- [37] ESA VEX Mission Definition Report 2001
- [38] Kuzmin, A.D. 1983, in “Venus”, Tucson, AZ, University of Arizona Press, 36
- [39] Gierasch, P.J., Goody, R.M., Young, R.E. et al. 1997, in “Venus II”, Tucson, AZ, University of Arizona Press, 459
- [40] Newman, M. and Leovy, C. 1992, *Science* 257, 647
- [41] Colin, L. 1983, in “Venus”, Tucson, AZ, University of Arizona Press 1983, 10
- [42] Clancy, R.T., Sandor, B.J., Moriarty-Schieven, G.H. 2003, *Icarus* 161, 1
- [43] Clancy, R.T., Sandor, B.J. and Moriarty-Schieven, G.H. 2003, *American Astronomical Society, DPS meeting* 39, 61.07
- [44] Seiff, A. 1983, in “Venus”, Tucson, AZ, University of Arizona Press 1983, 215
- [45] Titov, D.V., Svedhem, H., Koschny, D. et al. 2007, *PSS* 54, 1279

- [46] Lellouch, E., Clancy, T., Crisp, D. et al. 1997, in “Venus II”, Tucson, AZ, University of Arizona Press 1997, 295
- [47] Crisp, D. and Titov, D. 1997, in “Venus II”, Tucson, AZ, University of Arizona Press 1997, 295
- [48] Newman, M., Schubert, G., Kliore, A.J. et al. 1984, *J. Atmos. Sci.* 41, 1901
- [49] Zhang, S., Bougher, S.W. and Alexander, M.J. 1996, *JGR* 101, 23195
- [50] Brace, L.H., Taylor, H.A. Jr., Gombosi, T.I. et al. 1983, in “Venus”, Tucson, AZ, University of Arizona Press, 779
- [51] Bertaux, J.L., Vandaale, A.C., Korabev, O. et al. 2007, *Nature* 450, 646
- [52] Taylor, F.W., Hunten, D.M., Ksanfomality, L.V. 1983, in “Venus”, Tucson, AZ, University of Arizona Press, 650
- [53] Hüslér, B., Pätzold, M., Tyler, G.L. et al. 2006, *PSS* 54, 1315
- [54] Schubert, G. 1983, in “Venus”, Tucson, AZ, University of Arizona Press, 681
- [55] Sharma, R.D., Wintersteiner, P.P. 1990, *GRL* 17, 2201
- [56] Lopez-Puertas, M., Lopez-Valverde, M.A., Rinsland, C.P. et al. 1992, *JGR* 97, 20469
- [57] Roos-Serote, M., Drossart, P., Encrenaz, T. et al. 1995, *Icarus* 114, 300
- [58] Bougher, S.W., Rafkin, S., Drossart, P. 2007, *PSS* 54, 1371
- [59] Drossart, P. et al. 2007, *Nature* 450, 641, doi:10.1038/nature06140
- [60] Hinson, D.P., Jenkins, J.M. 1995, *Icarus* 114, 310
- [61] Nieman, H.B., Hartle, R.E. and Kasprzak, W.T. 1979, *Science* 203, 770
- [62] Mier, H.G., Harris, I., Kasprzak, W.T. et al. 1988, *JGR* 93, 11247
- [63] Forbes, J.M. and Konopliv, A. 2007, *GRL* 34, L08202, doi:10.1029/2007GL029252
- [64] Collard, A.D., Taylor, F.W. and Calcutt, S.B. 1993, *PSS* 41, 487
- [65] Lopez-Valverde, M.A., Lopez-Puertas, M., Mark, C.J. et al. 1993, *GRL* 20, 1247
- [66] Dickinson, R.E. 1972, *JAS* 47, 293
- [67] Deming, D. and Mumma, M.J. 1983, *Icarus* 55, 356
- [68] Gordiets, B.F. and Panchenko, V.I. 1983, *Cosmic Res.* 21, 725
- [69] Stepanova, G.I., Shved, G.M. 1985, *Sov. Astron.* 29, 422
- [70] Johnson, M.A., Betz, A.L., McLaren, R.A. et al. 1976, *ApJ* 208, L145
- [71] Deming, D., Espenak, F., Jennings, D. et al. 1983, *Icarus* 55, 347
- [72] McCleese, D.J., Haskins, R.D., Schofield, J.T. et al. 1992, *JGR* 97, 7735
- [73] Roldan, C., López-Valverde, M.A., López-Puertas, M. et al. 2000, *Icarus* 147, 11
- [74] Lopez-Valverde, M.A., Drossart, P., Carlson, R. et al. 2007, *PSS* 55, 1757
- [75] Gilli, G., Lopez-Valverde, M.A., Drossart, P. et al. 2008, *JGR* Submitted
- [76] Gerard, J., Saglam, A., Piccioni, G. et al. 2008, *GRL* in press
- [77] Donahue, T.M., Hoffman, J.H., Hodges, R.R. et al. 1982, *Science* 216, 630
- [78] Forget, F., Millour, E., Lebonnois, S. et al. 2006, 2nd workshop on Mars atmosphere modelling and observations, held February 27 - March 3, 2006 Granada, Spain
- [79] Kasting, J.F. and Pollack, J.B. 1981, *Icarus* 53, 479
- [80] Zhang, T.L., Delva, M., Baumjohann, W. et al. 2007, *Nature* 450, 654
- [81] Barabash, S., Fedorov, A., Sauvaud, J.J. et al. 2007, *Nature* 450, 650

# THE PHYSICAL CHARACTERIZATION OF THE STARS

CARME JORDI , EDUARD MASANA

*Departamento de Astronomía y Meteorología, Universidad de Barcelona, Martí i Franqués 1, E-08028 Barcelona, SPAIN*

*Institut de Ciències del Cosmos – Institut d’Estudis Espacials de Catalunya*

## **Abstract:**

The determination of the stellar parameters is of crucial importance for many fields of astrophysics, from the comprehension of the structure of the stars and their evolution, up to the determination of distances and the cosmological scale. This contribution reviews the main observational techniques and methods of calibration for the determination of the mass, radius, temperature, chemical composition and luminosity of stars.

**Keywords:** stars: chemical composition – stars: radius – stars: mass – stars: distance – stars: luminosity – stars: temperature

## 1 Introduction

The stars are the basic pieces of the universe and the understanding of their properties, structure, formation and evolution is one of the most important branches of astrophysics. To characterize the stars allows to characterize the galaxies and from them the universe as a whole. As it is well known, stars are formed in groups from the gravitational collapse of giant and cold clouds of molecular gas, take long episodes of production of energy through thermonuclear reactions in their interior with a length that depends on the mass, and die as highly compact objects. The life of an isolated star is mainly driven by their initial mass and chemical composition and along the years, it changes its observational properties, like size, luminosity, temperature, gravity and so on. Stars with the same initial properties have equivalent evolutions, and at the same evolutionary stage they have equal intrinsic luminosities. They may be seen with different apparent luminosities due to the attenuation by the distance and the extinction. This way, the comparison of observed luminosities of different stars (assumed to be intrinsically similar<sup>1</sup>) is a way to determine stellar distances, and actually it is the basis of the cosmic distance scale.

---

<sup>1</sup>The similarity of stars is derived from comparison of their spectral types, photometric colour indices, strenght of spectral lines, etc

Therefore, it is extremely important to characterize physically all of the stars, or in other words, it is important to determine the physical fundamental parameters (mass, radius, temperature, luminosity, composition).

Unlike other experimental sciences, astronomy and astrophysics cannot perform laboratory tests with stars. The astronomers must analyze the electromagnetic radiation coming from the star to deduce all the physical information. The most frequent observational methodologies, photometry and spectroscopy, allow to derive the flux measured by the observer in different wavelength ranges. The detailed analysis of this information allows to deduce the observable parameters of the stars: the temperature, gravity, luminosity and composition. The variation with time of the measured flux and of its wavelength distribution can be indicative of the presence of companion stars. In some of these cases, it is possible to determine the masses and the radii of the involved stars.

This contribution describes the principal methods that allow the determination of the physical stellar parameters, the current state of the knowledge and the future perspectives. We do not pretend to be complete, but to list the most used approaches.

## 2 Chemical composition

The chemical composition is measured from the intensities of the lines and bands of absorption. Here the most precise approach is to use high resolution spectroscopy. The fitting of synthetic spectral energy distributions (SED), computed with stellar atmosphere models, to the observed spectra provide the chemical abundances. To do that, effective temperatures, surface gravities and rotational velocities have to be assumed.

Among the catalogues of spectroscopic abundances, we mention the compilation by [9] that includes 5946 measurements of 3247 stars. A later edition by [10] contains [Fe/H] for FGK stars derived from high resolution and high S/N spectra. The catalogue by [14] contains detailed abundances of elements, including the  $\alpha$ -elements, for 189 nearby field F and G dwarfs. Standard deviations of measurements are usually at the level 0.05–0.10 dex, and accuracies are at the level of 0.10–0.15 dex.

Spectroscopy needs long exposure times with large telescopes and in the case of faint stars or for large surveys, this can turn out to be highly prohibitive. The photometric measurements with intermediate bands located at suitable spectral ranges is an alternative and competitive approach, provided that the changes of the spectra due to changes of chemical composition have been calibrated through spectroscopic measurements ‘a priori’. Several photometric systems have been designed for such a purpose, and among them we would like to mention the *uvby* Strömrgren system [37], the Vilnius system [36] and the Gaia system [18].

Figure 1, extracted from [21], shows the comparison of spectroscopic and photo-

metric determinations of  $[\text{Fe}/\text{H}]$ . The photometric derivation uses the *uvby* system and the calibration of [35]. In this case, the photometric calibration yields abundances slightly higher than the spectroscopic ones. Once corrected, the precisions are of the order of 0.15-0.20 dex.

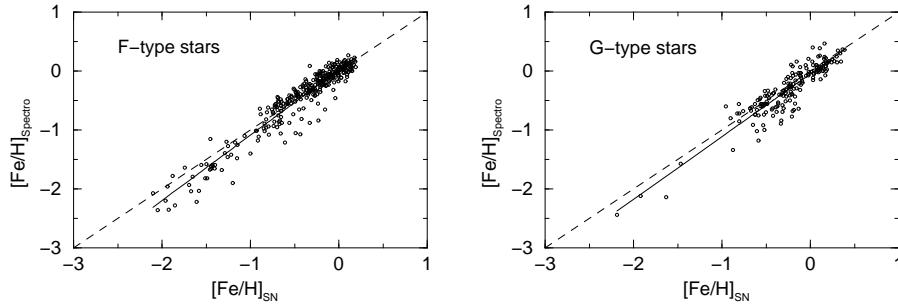


Figure 1: Comparison of the determination of chemical compositions from photometry and spectroscopy (extracted from [21]). The photometric system is the *uvby* system and the calibration is that of [35].

Guidelines for designing a photometric system sensitive to C,N,O and  $\alpha$ -process elements were provided by [38] and [39]. The CaII H and K lines and MgI b triplet are the most sensitive direct indicators of  $[\alpha/\text{Fe}]$  changes. Figure 2 (from [38]) shows the variation of the spectra when the effective temperature or the surface gravity or the chemical abundances change. The authors conclude that narrow bands placed on the Ca II H and K lines and Mg Ib triplet would allow to disentangle the changes due to variations of Fe- and  $\alpha$ -elements. This idea was incorporated into the Gaia system [18] design.

### 3 Radius

Two techniques of observations are used to measure stellar angular diameters: the high angular resolution interferometry and the lunar occultations. While the first is a direct measure of the diameter, the second is a measure of the time needed for the star to be hidden rear the Moon, which is related with the angular diameter of the star and the motion of the Moon and the observer. The CHARM2 catalogue ([32]) is a compilation of direct measurements by high angular resolution methods, as well as indirect estimates of stellar diameters. A total of 8231 entries for 3238 unique sources are present.

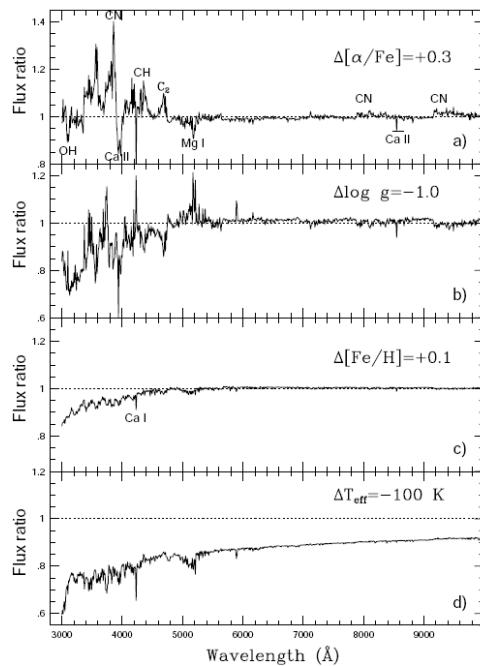


Figure 2: Relative variation of the flux energy with changes of effective temperature, surface gravity or chemical abundances (extracted from [38])

Other methods to derive angular diameters are based on the comparison of observed and theoretical spectral energy distribution. See [34] for a discussion, and [22] for an example. Figure 3 shows a comparison of semi-diameters derived by [22] with the above empirical determinations in the CHARMM2 catalogue considering an uniform disk. The agreement is excellent. A crude comparison of both uniform disk and limb darkened values for about 1600 F, G and K stars indicates a  $\sim 4\%$  positive correction for limb darkening, of the same order of the dispersion shown in Figure 3.

Double-lined eclipsing binaries ([27]) and planetary transits ([8]) yield absolute stellar radii. Actually, radii relative to the semi-major axis of the orbit are derived from the light curves, and the inclination and the radial velocity curves complete the needed data to obtain the absolute radii. The light curves analysis takes into account corrections by limb darkening. The double-lined eclipsing binaries provide the most accurate determinations at present, which are  $\sim 1\text{--}5\%$ .

The Baade-Wesselink method applied to expanding or pulsating photospheres is

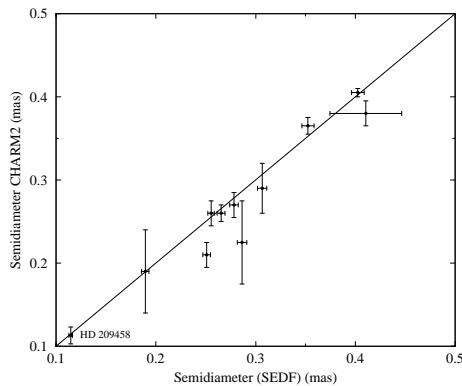


Figure 3: Comparison of angular semi-diameters computed from Spectral Energy Distribution Fit (SEDF) method [22] and from CHARMM2 catalogue [32]. In the case of HD 209458, the comparison of the semi-diameter is between SEDF method and an empirical determination from a high-precision transit light curve [8]. (Extracted from [22])

discussed by [34]. Under these special circumstances observed Doppler shifts of lines originating in moving photospheres may be used to derive absolute stellar diameters.

We would like to remark that the combination of temperatures and absolute radii allows to derive luminosities and, hence, direct distances with precisions almost equivalent to trigonometric parallaxes. This method has already allowed a precise determination (2-5%) of the distance of individual stars in LMC ([16] among others) and M31 [31]. The era of big telescopes will allow to extend substantially the sample and to reduce the current uncertainty.

## 4 Mass

The mass of a star can only be directly known through the analysis of its gravitational effect on the neighbouring objects, either stars or planets. In many cases, the multiple system is not resolved and it is discovered by periodic Doppler shifts of the spectral lines or periodic fluctuations of the apparent magnitude due to eclipses, because of the revolving motion of the star around the center of mass of the system.

In the case of double-lined spectroscopic binaries, the two stars have similar luminosity and the lines of both are distinguishable in the spectrum. This allows to derive the ratio of masses as well as the semi-major axis of the orbit by the factor  $\sin i$ , being  $i$  the inclination of the orbit with respect the line of sight. If eclipses occur,

the analysis of the light curve yields the inclination (among other parameters), and therefore the absolute individual masses can be derived.

The double-lined eclipsing binaries are extremely important, since they are the only case providing simultaneous determinations of individual masses and radii (see Sect. 3). The best reached precisions are of the order of 1-5% ([27], [4]). Such precisions are really needed for detailed tests of stellar structure and evolution models. Several tests have revealed discrepancies at both ends of the main sequence (high and low mass, see for instance [29] and [23]).

The masses of single stars can be estimated through Mass-Luminosity relationships based on binaries (see for instance [27, 4]), or through theoretical models. Theoretical stellar evolutionary models yield the temperature and luminosity (or radius) of a star of a given mass and chemical composition at a certain age. The location of a star in the observational colour-magnitude diagram can be transformed into a location in the theoretical luminosity-temperature diagram and from it estimate the mass and the age using the evolutionary tracks. Theoretical and observable quantities can be related with the use of appropriate calibrations and stellar atmosphere models. To cite one example, Asiain et al [5] derived stellar masses and ages for main sequence A-type stars with metallicity, effective temperature and surface gravity derived from  $wby - \beta$  photometry.

Microensing phenomena can provide direct determinations of masses of single stars. When a star is acting as a lens, it is a non-massive lens that yields two unresolved images of the foreground object. The photocenter position of the composed image and its brightness change with the alignment, and the duration of the transient phenomena depends on the relative proper motions of the lens and the foreground object. The astrometric analysis of the photocenter displacement yields a direct measurement of the mass of the lens. The displacement is of a milli-arcsecond or less and hence an astrometric precision as small as few tens of micro-arcseconds is needed. VERA (VLBI Exploration of Radio Astrometry) [17] and the next Gaia ESA's mission [25, 20, 26] are able to provide such precisions and so masses of single field stars.

## 5 Effective temperature

There are several approaches in the literature to compute effective temperatures. Except when applied to the Sun, very few of them are *direct* methods that permit an empirical measurement. Usually, *semi-empiric* or *indirect* methods are based to a certain extent on stellar atmosphere models. Among the *direct* approaches we find the remarkable work by [12], which is based on interferometric measurements of stellar angular semi-diameters  $\theta$  and total fluxes  $F_{\text{bol}}$  at Earth ( $F_{\text{bol}} = \theta^2 \sigma T_{\text{eff}}^4$ ). This work contains data for 32 stars of O5-F8 spectral types. Although being the most direct

method for temperature determination, it still needs to rely at some extent on stellar atmosphere models to predict the energy in the wavelength range not covered by the observations and to correct the semi-diameters measurements for limb darkening.

The *indirect* methods are mainly based on the use of photometry, spectroscopy, or a combination of both. In photometry, the technique is to measure the radiation in the pseudo-continuum of the spectra, providing color indices free from chemical abundance signatures and so related with temperature (and interstellar absorption). As often as possible it is necessary to look for reddening-free indicators, like the intensity of certain spectral lines or bands ( $H_\beta$  line, TiO molecular bands, for instance), IR indices, etc. Or, either derive the interstellar absorption independently and correct the pseudo-continuum color indices. The synthetic photometry of SED predicted by atmosphere models, or the photometry of stars in [12] provide the necessary calibrations to relate photometric indices with temperatures. A good review of the problems involved in such temperature – color relations is provided by [6].

The use of IR photometry to determine effective temperatures was initially proposed by [7]. Their so-called Infrared Flux Method (IRFM) uses the ratio between the bolometric flux of the star and the monochromatic flux at a given infrared wavelength, both measured at Earth, as the observable quantity. This ratio is then compared with a theoretical estimate derived from stellar atmosphere models to carry out the determination of the effective temperature. The IRFM has been widely used by several authors, being most noteworthy the work by [1, 2, 3].

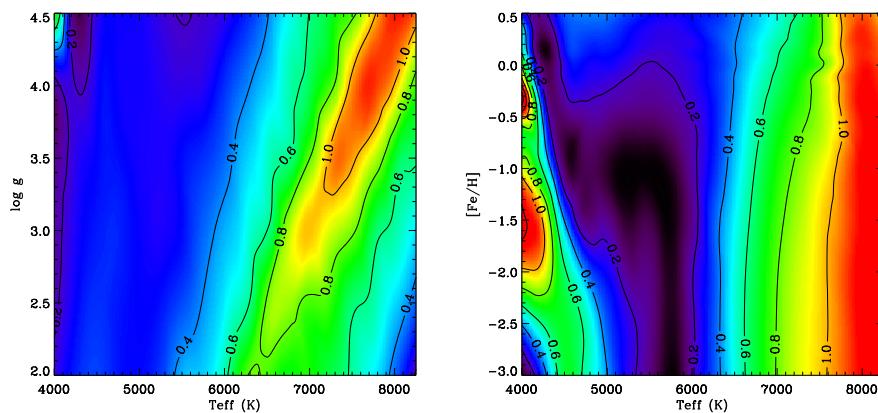


Figure 4: Relative error in temperature due to an uncertainty of 0.5 dex in gravity, derived using the Spectral Energy Distribution Fir method. Left: for solar metallicity. Right: for gravity 4.5. (extracted from [21])

The Spectral Energy Distribution Fit (SEDF) method proposed by [22] follows a somewhat different approach, namely the fit of the stellar spectral energy distribution from the optical ( $V$ ) to the IR ( $JHK$ ) using synthetic photometry computed from stellar atmosphere models. Unlike the implementation of the IRFM by [2], which averages temperatures derived individually for each IR band, the SEDF method takes into account the four bands simultaneously. In addition, and also unlike the IRFM, the bolometric flux is not required *a priori* but results self-consistently with the temperature determination. A fitting algorithm minimizes the difference between observed and synthetic photometry by tuning the values of the effective temperature and the angular semi-diameter. In addition, the  $BC$  can be obtained from these two parameters, and then, when the distance to the star is known, the luminosity is computed from the  $BC$  and the absolute magnitude in a given photometric band. The uncertainties of the derived parameters (temperatures, angular semi-diameter and  $BC$ ) are estimated from the errors in the observed and synthetic photometry as well as in the assumed  $[Fe/H]$ ,  $\log g$  and  $A_V$ . The relative error in temperature for several uncertainties of  $[Fe/H]$ ,  $\log g$  and  $A_V$  are evaluated in [22]. As an example, Fig. 4 shows the relative error in temperature due to an uncertainty of 0.5 dex in gravity. For a sample of about 11 000 Hipparcos stars, and using 2MASS photometry ([13]), [22] derive temperatures with mean precisions of about 1%. Through an extensive comparison with determinations by other authors, [22] conclude that the effective temperature scale of FGK stars (4000–8000 K) is currently established with a net accuracy better than 0.5–1.0%.

Figure 5 shows the calibration by [22] of effective temperature as a function of the color index  $V - K_S$  and the chemical composition.

In [28], direct temperatures for a sample of double-lined eclipsing binaries with accurate radii (1-2%) and relative error in the Hipparcos parallax below 20% are derived. The temperature is obtained from the Hipparcos parallax, the radius of the star, the visual magnitude and the bolometric correction, and thus, it is a rather direct method, since it only needs calibrations for the BC value. The comparison with photometric determinations of temperatures shows a small trend leading to the parallax based determination about 2-3% smaller than the photometric one. Due to the completely independent nature of the two temperature determinations, the authors conclude that a small systematic difference of about 0.012 dex is present in the temperature range covered by their sample (from 5000 K to 25000 K). Moreover, the two binary systems of the sample with the smallest relative errors in the Hipparcos parallax,  $\beta$  Aur and V1143 Cyg, agree with the photometric data. Highly accurate trigonometric parallaxes from the next Gaia ESA's mission [25, 20] will yield very accurate temperatures using this approach.

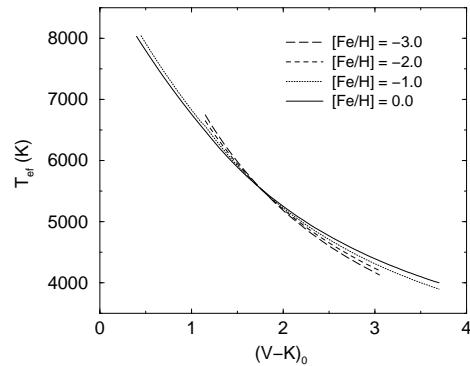


Figure 5: Relation between effective temperature and the color index  $V - K_S$  for FGK dwarfs. The dependence with the chemical composition is shown as well. Figure extracted from [22]

## 6 Luminosity and distance

The absolute magnitude (and/or the luminosity) is, probably, the stellar parameter with the greatest cosmological transcendence. The direct measure of the intrinsic luminosity of the star is not possible. Only it can be realized if the effect of the interstellar absorption as well as the distance are known. The most precise measurement of the total energy released by the star is the combination of ground- and space-based observations with stellar atmosphere models to predict the energy in the wavelength range not covered by the observations.

Many absolute magnitude calibrations as a function of broad- or medium-band color indices or spectral types exists (see for instance [33]). All of them rely on primary calibrators with well known stellar distances. The accuracies of those calibrations range from 10 to 30% depending on the indicators used.

The large stellar distances make the measure of the trigonometric parallaxes only possible for the most nearby stars. The most accurate parallaxes nowadays available come from the Hipparcos mission [15]. Hipparcos constituted a great step forward in the measure of parallaxes, yielding  $\sim 21\,000$  stars with relative errors lower than 10% and  $\sim 50\,000$  stars with relative errors lower than 20%. A complete review of the scientific exploitation of the Hipparcos data and specially the derivation of distances and luminosities can be found in [24]. The current knowledge of luminosities of the many types of stars across the HR diagram is discussed. The extremely precise parallaxes force the use of statistical methods to account for the biases in the observational samples. The next mission Gaia, to be launched at the end of 2011,

will provide extremely precise parallaxes for some million stars including stars of all types.

## 7 Summary

We have listed several observational techniques and calibration methods to derive the fundamental properties of the stars. The most used are summarized as follows:

- High resolution spectroscopy yields the determination of chemical composition and allows the calibration of photometric indices sensitive to changes of the elements abundances
- High angular resolution interferometry and lunar occultations yield angular semi-diameters
- Comparison of observed and theoretical spectral energy distributions yield temperatures and semi-diameters. If distances are known, absolute radii can be derived
- Unreddened or reddening free color indices provide temperatures through suitable calibrations
- Photometric and spectroscopic observations of double-lined eclipsing binaries allow the simultaneous determination of individual masses and radii (at the level of 1-5% accuracy). Moreover, if the distance is known, the temperature can be derived. Instead, if the temperature is known, the distance can be computed
- Masses of individual stars can be estimated with a Mass-Luminosity relation or through stellar structure models
- Absolute magnitudes and luminosities can be determined ‘directly’ if parallaxes are known, or ‘indirectly’ through calibrations primarily based on a set of closer-by stars.

*Acknowledgements:* This work has been supported by the Spanish MCyT AYA2006-15623-C02-02 project.

## References

- [1] Alonso, A., Arribas, S., Martínez-Roger, C. 1995, A&A 297, 197
- [2] Alonso, A., Arribas, S., Martínez-Roger, C. 1995, A&AS 117, 227
- [3] Alonso, A., Arribas, S., Martínez-Roger, C. 1995, A&A 313, 873
- [4] Andersen, J. 1991 A&ARv 3, 91
- [5] Asiain, R., Torra, J., Figueras, F. 1997, A&A 322, 147
- [6] Bell, R.A. 1997, in 'Fundamental Stellar Properties: The interaction between observations and theory' (Bedding et al. eds.), 159
- [7] Blackwell, D. E. and Shallis, M. J. 1977, MNRAS 180, 177
- [8] Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., Burrows, A. 2001, ApJ 552, 699
- [9] Cayrel de Strobel, G., Soubiran, C., Friel, E., Ralite, N., Francois, P. 1997, A&AS 124, 299
- [10] Cayrel de Strobel, G., Soubiran, C., Ralite, N. 2001, A&A 159, 373
- [11] Crawford, D.L. 1960, AJ 65, 487
- [12] Code, A.D., David, J., Bless, R.C., Hanbury-Brown, R. 1976, ApJ 203, 417
- [13] Cutri, R.M., Skrutskie, M.F., van Dyk, s., et al. 2003, <http://www.ipac.caltech.edu/2mass/releases/allsky/doc/explsup.html>
- [14] Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A 275, 101
- [15] ESA, 1997, The Hipparcos and Tycho Catalogues, ESA SP-1200
- [16] Guinan, E. F., Fitzpatrick, E. L., Dewarf, L. E. et al 1998, ApJL 509, 21
- [17] Honma, M. 2001, PASP 53, 233
- [18] Jordi, C., Høeg, E., Brown, A.G.A. et al. 2006, MNRAS 367, 290
- [19] Jordi, C., Luri, X., Masana, E., et al. 2002, Highlights of Astronomy, Vol. 12, 684
- [20] Lindegren, L., Babusiaux, C., Bailer-Jones, C. et al. 2008, Procs. of the IAU Symp 248, A Giant Step: From Milli- to Micro-arcsecond Astrometry, W.J. Jin, I. Platais & M.A.C. Perryman, eds., in press
- [21] Masana, E. 2004, Tesis doctoral, Universidad de Barcelona
- [22] Masana, E., Jordi, C., Ribas, I. 2006, A&A 450, 735
- [23] Morales, J.C., Ribas, I., Jordi, C. 2008 A&A in press arXiv:0711.3523
- [24] Perryman, M.A.C. 2008, Astronomical Applications of Astrometry: A Review Based on Ten Years of Exploitation of the Hipparcos Satellite Data, Cambridge University Press (in press)
- [25] Perryman, M.A.C., de Boer, K.S., Gilmore, G. et al. 2001, A&A 369, 339
- [26] Perryman, M.A.C., de Boer, K.S., Gilmore, G. et al. 2000, GAIA, Composition, Formation and Evolution of the Galaxy, Concept and Technology Study Report, ESA-SCI(2000)4
- [27] Popper, D.M. 1980 ARA&A 18, 115
- [28] Ribas, I., Giménez, A., Torra, J., Jordi, C., Oblak, E. 1998, A&A 330, 600
- [29] Ribas, I., Jordi, C., Giménez, A. 2000, MNRAS 318, 55
- [30] Ribas, I., Jordi, C., Torra, J. 1999, MNRAS 309, 199

- [31] Ribas, I., Jordi, C., Vilardell, F., et al. 2005, ApJL 635, 37
- [32] Richichi, A., Percheron, I., Khristoforova, M. 2005, A&A 431, 773
- [33] Schmidt-Kaler, T., 1982, in Landolt-Börnstein, eds. K. Schaifers and H.H. Voigt, Vol. II, Subvol. B, 453
- [34] Scholz, M., 1997, in 'Fundamental Stellar Properties: The interaction between observations and theory' (Bedding et al. eds.), 51
- [35] Schuster, W.J., Nissen, P.E. 1989, A&A 221, 65
- [36] Straizys, V., Zdanavičius, K. 1965, Bull. Vilnius Obs. 14, 1
- [37] Strömgren, B. 1966, ARA&A 4, 433
- [38] Tautvaišienė G., Edvardsson B. 2002, Ap&SS, 280, 143
- [39] Tautvaišienė G., Edvardsson B., Bartašiutė S. 2003, Baltic Astronomy 12, 532

# DARK ENERGY AND THERMONUCLEAR SUPERNOVAE

INMA DOMÍNGUEZ

*Dpto. de Física Teórica y del Cosmos, Facultad de Ciencias, Universidad de Granada, E-18071 Granada, SPAIN*

EDUARDO BRAVO

*Dpto. de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, SPAIN*

LUCIANO PIERSANTI , AMEDEO TORNAMBE , OSCAR STRANIERO

*Osservatorio Astronomico di Collurania-INAf, Via Mentore Maggini 47, 64100 Teramo, ITALY*

PETER HÖFLICH

*Dept. of Physics, Florida State University, 315 Keen Building Tallahassee, FL 32306-4350, USA*

## **Abstract:**

Nowadays it is widely accepted that the current Universe is dominated by dark energy and exotic matter, the so called Standard Model of Cosmology or  $\Lambda$ CDM model. All the available data (Thermonuclear Supernovae, Cosmic Microwave Background, Baryon Acoustic Oscillations, Large Scale Structure, etc.) are compatible with a flat Universe made by  $\sim 70\%$  of dark energy. Up to now observations agree that dark energy may be the vacuum energy (or cosmological constant) although improvements are needed to constrain further its equation of state. In this context, the cosmic destiny of the Universe is no longer linked to its geometry but to the nature of dark energy; it may be flat and expand forever or collapse. To understand the nature of dark energy is probably the most fundamental problem in physics today; it may open new roads of knowledge and led to unify gravity with the other fundamental interactions in nature. It is expected that astronomical data will continue to provide directions to theorists and experimental physicists. Type Ia supernovae (SNe Ia) have played a fundamental role, showing the acceleration of the expansion rate of the Universe a decade ago, and up to now they are the only astronomical observations that provide a direct evidence of the acceleration. However, in order to determine the source of the dark energy term it is mandatory to improve the precision of supernovae as distance indicators on cosmological scale.

**Keywords:** Stellar Evolution Models – White Dwarfs – Supernovae – Cosmology.

## 1 Introduction

Two papers, published a decade ago, marked for the scientific community the beginning of the accelerating Universe: *Observational evidence from supernovae for an accelerating universe and cosmological constant* (September, 1998 [1, 2]) by the High-z Team and *Measurements of  $\Omega$  and  $\Lambda$  from 42 high-redshifts supernovae* (June 1999 [3]) by the Supernova Cosmology Project. The first one is based on the data of 10 SNe up to  $z=0.62$  and the second one on the data of 42 SNe up to  $z=0.83$ .

These two independent teams got the same result: acceleration of the expansion and, consequently, a nonzero  $\Lambda$ . Note that they were trying to measure the opposite, the deceleration of the expansion, and, moreover, this surprising result was based on the fact that the objects were dimmer than expected; it was unlikely that for any unknown reason less luminous objects were by chance observed instead of the brighter ones.

This exotic component was the missing piece in the Universe component puzzle and finally several long standing problems converged to a solution. *The Cosmological constant is back*, was the title of a paper published by Kraus & Turner in 1995 [4], 3 years before the SNe Ia results came out. Since the beginning, the analysis of Cosmic Microwave Background radiation [5, 6, 7] favoured a flat geometry and other independent set of observations, like the large scale structure of the Universe, favoured a low density-matter Universe, compatible with 30% matter [8]. All three experiments are in perfect agreement and in order to have  $\Omega_\lambda=0$ , two of them would have to contain severe systematic errors [9].

Additionally, as an accelerating Universe is older ( $\sim 14500$  Ma assuming a Hubble constant of 65 km/s/Mpc), the old problem with respect to the age of the oldest globular clusters is *nearly* solved (in our opinion globular cluster would better fix in a *slightly* older Universe).

Evidence for cosmic acceleration has gotten stronger in all these years and nowadays, all efforts focus on characterizing the dark energy equation of state (EOS). The simplest equation relates pressure and density by a parameter,  $P = w\rho$ ;  $w$  may be a constant ( $w=-1$  for the Cosmological constant or vacuum energy) or vary with redshift. Observations have not provided any evidence that dark energy is not the energy related to the quantum vacuum ( $w=-1$  and constant) although theory does not explain it. Empty space is not really empty, particle-antiparticle pairs appear and disappear existing for very brief time but current estimation of the corresponding vacuum energy is  $10^{120}$  times more than needed to explain the acceleration. It is probably the biggest discrepancy between theory and observations in all physics. Matter dilutes as the Universe expands while the vacuum energy remains constant; hence, the matter term dominated in the past while vacuum energy will completely dominate in the future. So, if vacuum energy is really the dark energy, why its density is now so close to that of matter? It is important to understand better the energy of

*nothing.*

A strong indication in favour of  $w = -1$  has come from SNIa Hubble diagrams, which currently includes 200 SNe up to  $z = 1.8$ ; in fact, they show the expected deceleration at  $z > 0.5$ , and recent acceleration [10].

The future is promising: several experiments, aimed to elucidate the properties of dark energy, are underway or planned from ground based telescopes, from the space, accelerators, and underground laboratories.

Note that the only direct-direct evidence of the acceleration of the expansion rate of the Universe is provided by SNIa Hubble diagrams. It is worth to extend the observations to  $z > 1$  where the early effects of deceleration may be detectable [10] and to improve supernova distances and analyze their limits for precision measures of dark energy.

## 2 Thermonuclear supernova Hubble diagrams: A component with negative pressure is needed

SNIa Hubble diagrams have been used for cosmological applications since 1968 [11]; at that time the dispersion was 0.6 mag. Later on, core-collapse events were identified (Ib) and Hubble diagrams, based only on Type Ia, improved. At the beginning, SNe Ia were considered to be very similar, if not identical, each other, but it became soon evident that this hypothesis was not correct. In 1991, the discovery of the superluminous 1991T and the subluminous 1991bg pointed out that Type Ia were not standard candles.

In 1993 Phillips [12] showed that for a number of near and well observed SNe Ia, whose distance was *a-priori* determined by means of *bona fide* distance indicators, there exists a linear correlation between maximum luminosity and decline rate of luminosity after maximum: peak luminosity decreasing with an increasing decline rate. Basing on this evidence, these Supernovae can be considered once again as an homogeneous class and they can be used as *calibrated candles* using the maximum-decline relation or equivalents [12, 13, 14, 15, 16]. These relations have been improved, mainly by extinction corrections, up to a dispersion of  $\sigma \leq 0.18$ .

At the same time, the technics to discover high- $z$  SNe developed [17] and *big* CCDs were constructed, like the *Big Throughput Camera* by Tyson and Bernstein that became available for the community at Cerro Tololo and was soon used by the two teams, High- $z$  Team and SN Cosmology Project. An amazing result came out and, since then, our understanding of the Universe completely changed. The conclusion was that SNe Ia located at  $0.4 \leq z \leq 0.9$  were about 0.2 magnitudes dimmer than nearby SNe, meaning that they were located further than expected for a constant expansion rate of the Universe (an empty Universe). A negative pressure was required to balance the gravity effects and accelerate the expansion.

Most important, the observations of Type Ia supernovae by themselves indicate the existence of dark energy [18] and show deceleration at  $z \geq 0.5$ , and acceleration at lower  $z$ , ruling out pure acceleration and gray-dust absorption, all at 99% confidence level. Note that these results are obtained with a dispersion in the calibration relation of the order of  $\sigma \leq 0.2$  magnitudes.

## 2.1 EOS of dark energy: A precision better than 0.02 mag is required

Present and future experiments focus on better characterizing the equation of state (EOS) of dark energy. When SNe Ia are used to characterize  $w$ , two constraints are usually assumed, a flat ( $\Omega_{tot} = 1$ ) and low-matter density ( $\Omega_m = 0.3$ ) Universe. Notwithstanding these assumptions, it is necessary to have an accuracy of better than 0.02 magnitudes to discriminate between Universes with different  $w$  and it is clear that systematic effects will limit our progress.

It is noteworthy that, without the a priori constraint of a flat Universe, the cosmological parameters that best fit Type Ia SNe measurements may be  $\Omega_\lambda=1.57$  and  $\Omega_m=0.79$  [19].

Now it is mandatory to reduce the current scatter in the calibration relation by a factor of 10, otherwise we may have already reached SNIa limits to measure Cosmology.

## 3 Systematic effects

The reliability of the maximum-decline relation when it is applied outside the local sample depends on the identification of systematic effects. The high redshift sample could be contaminated with peculiar Ia and with core collapse Ib/c SNe. Some nearby underluminous SNe, like 2002cx [20], 2005hk [21] and 2005bl [22] are well outside the maximum-decline relation although in these cases spectra, taken at the epoch of maximum light, easily reveal their different origin.

However, this approach does not always work; SN2003fg at  $z = 0.2$  [23] appears normal at all levels, except for the high luminosity, and it is well outside the maximum-decline relation. In fact, a super-Chandrasehar white dwarf (WD) has been proposed as its progenitor.

Other sources of uncertainty could be K-corrections, the correction that has to be applied to account for the differences caused by the spectrum shifting; the light of the SN is shifted to longer wavelengths while astronomical observations are made in fixed band passes on Earth (e.g., R-band matched to B-band at  $z = 0.5$ ). The error due to uncertainties in this correction is estimated to be small,  $\sim 0.01$  mag [9].

Uncertainties of major concern are those related with potential evolutionary effects: dependence of the SN properties on redshift, changes in the extinction law and, if SN are sensitive to their progenitors, changes in the stellar populations from which they originate.

In the local Universe we see SNe Ia in a variety of environments, some of them with significant extinction and part of them with *peculiar* extinction laws [24]. The properties of the intervening dust may be different going back in time; e.g., a smaller contribution from low mass red giants due to their longer evolutionary times. As the number of SNe Ia increases it is possible to construct Hubble diagrams with subsamples; e.g., only SNe occurring in passive galaxies, less affected by absorption. Most interesting, it has been shown that SNe Ia are good standard candles in the IR; at these wavelengths the extinction is smaller and SNe Hubble diagrams in the IR are promising [25].

## 4 Observed dependence of SNIa properties on their host galaxies

Great efforts have been devoted in recent years to elucidate the observed dependence of the SN properties on the populations from which they originate. The first studies [26, 27, 14, 28, 29] revealed a dependence of SN properties on their host galaxies: slower light curves (LCs) (brighter SNe) do not occur in ellipticals, while the number of Type Ia SNe, per unit mass, is higher in late type galaxies than in ellipticals [30, 31] and, in star forming galaxies, the SN rate is higher in disks than in bulges [32].

Recently, Mannucci et al. [33] reported that the SNIa rate in late type galaxies is a factor  $\sim 20$  higher than in E/S0. Sullivan et al. [34] focused on the high-redshift sample and qualitatively derived similar trends to those obtained previously for the local Universe: (1) rates are higher (10 times) for strong star-forming galaxies than for galaxies without star formation, (2) slow-decline (bright) SN are only hosted by star-forming galaxies, while non-star-forming galaxies only host fast-decline (dim) SNe and (3) none of the dimmer SNe are found at high-redshift. Except for the most underluminous SNe, the distribution is identical for the high and the low-redshift samples.

These results imply that SNe come from at least two different stellar progenitors, old and young, that differ in 2-3 Gyr, and that their LCs properties are different, depending on those progenitors. At this point, we should note that the empirical calibration used in Cosmology is based on SNe coming from all types of galaxies and thus, the differences due to stellar populations in the nearby Universe are already within the brightness-decline relation. At the other hand, the relative contribution of the two components is expected to evolve strongly with time [35, 34]. A good understanding of the calibration relation in different environments will be essential

to fully exploit SNe Ia in measurements of  $w$ .

## 5 Our understanding of thermonuclear SNe

### 5.1 What we understand

Type Ia SNe are considered to be the thermonuclear explosion of carbon-oxygen (CO) WDs with a mass close to the Chandrasekhar mass [36]. Starting from such a structure and with a proper explosion mechanism, most of the observed properties are reproduced. The LC is powered by the radioactive energy coming from the decay of  $^{56}\text{Ni}$  to  $^{56}\text{Co}$  and  $^{56}\text{Co}$  to  $^{56}\text{Fe}$  and the maximum luminosity depends mainly on the  $^{56}\text{Ni}$  mass.

In fact, we have learnt a lot from observations and from one dimensional (1D) numerical simulations. Little carbon is observed in the spectra, indicating that nearly all the WD is burnt, which implies that the explosion (nuclear) energy is similar in all events and excludes a pure deflagration (subsonic burning front) that would leave an observable amount of matter unburnt in the outer part (see [37, 38] and references therein). Final kinetic energies are also similar, as the binding energy, which depends on the total mass and central density of the WD, is not expected to be very different. Intermediate mass elements (e.g. Si, Ca, Mg) are observed in the spectra around maximum light, implying that the external layers have experienced incomplete burning. This excludes a pure detonation (supersonic burning front) that would burn all the WD to Ni and others Fe-peak elements. The delayed detonation models [39, 40], in which the burning front is initially subsonic and at some point accelerates to supersonic velocities, fulfil the observational constraints. In these models, the inner part is completely burnt during the first deflagration phase while the external layers expand, the density decreases, and the right nucleosynthesis, a combination of complete and incomplete burning depending on density, is obtained. Moreover, if the transition from a deflagration to a detonation is delayed (or anticipated), the amount of  $^{56}\text{Ni}$  decreases (or increases) and the observed range of luminosities at maximum is reproduced.

Many studies have shown that the delayed detonation explosion of a Chandrasekhar mass ( $M_{Ch}$ ) WD satisfactorily explains most of the observed properties: the evolution of the LCs and spectra in the optical and in the IR, and also, the observed correlations, including the crucial maximum decline relation [41, 42, 43, 44, 45, 46, 47, 38].

For the reasons above, we have adopted this explosion mechanism and we have kept constant the needed parameters in the description of the deflagration velocity and the critical transition density: the density at which the change from a deflagration to a detonation takes place.

## 5.2 What we do not understand

The basic ideas explaining SNe Ia as the thermonuclear explosion of a CO WD with a mass close to  $M_{Ch}$  were established half a century ago but, in spite of all the dedicated resources, two reasons still prevent our full understanding of these events: (a) the nature and evolutionary history of the progenitor stellar system and (b) the characteristics of the explosion.

Single stellar evolution provides CO WDs with a maximum mass of around  $1.1 M_{\odot}$ ; when a CO core with a higher mass is formed, C is ignited [48]. Binary systems give the possibility of accreting mass from the companion and also explain that this explosions occur in passive galaxies where star formation ended several Gyrs ago. However, how a typical WD of  $0.6$  to  $1 M_{\odot}$  could reach the Chandrasekhar mass by accretion is still a matter of hot debate. Accretion of H or He from a non-degenerate companion is referred as the Single Degenerate (SD) scenario while accretion of C and O, from a thick disk formed from the merging of two WDs, is referred as the Double Degenerate (DD) scenario.

In the first case the donor is a normal star with an Hydrogen rich envelope [49]. These systems are well studied and they are associated with a variety of observational events (symbiotic stars, classical novae, cataclysmic variables, dwarf novae, recurrent novae and many others). However, in our opinion this scenario is not promising as SNe Ia progenitor since during the accretion process some instabilities arise, thus preventing the growth in mass of the CO core up to the Chandrasekhar limit [51]. For a different view see [50] and references therein.

The second scenario envisions a system composed by two tidally interacting WDs with total mass of the order of or greater than the  $M_{Ch}$ . In this case the two stars, due to gravitational wave radiation (GWR) emission, undergo a merging so that the less massive component completely disrupts, thus forming an accretion disk around the more massive one [53, 54]. In this case C-O rich matter is directly accreted. One major concern afflicts the DD scenario: the accretion rate which results after the merging would be so high that off-center carbon ignition occurs well before the WD could attain the limiting mass, thus producing as final outcome not a SNe Ia event, but an O-Ne white dwarf which eventually collapses into a neutron star [55, 56, 57]. The historical concern about the existence or not of observed counterparts of these systems seems to have been solved in favor of their existence [58]. Later on, we will show that, when the rotational effects are taken into account in modeling the binary system, the evolution from the merging of the two CO WDs up to the final explosive outcome is not only possible, but it results mandatory. Moreover, recent studies of observational properties of known Type Ia supernova remnants in the Galaxy favour the DD scenario [59].

Concerning the explosive phase, once C is ignited, the burning conditions are not reached by the compression of subsequent shells by the pressure wave (detonation). If

so, all the WD would be transformed to Fe-peak elements and no intermediate mass elements would be produced. On the other hand, if subsequent zones are heated only via thermal conduction, the velocity of the flame would be too low and the pressure wave would expand the WD till low densities, quenching the flame. However, turbulent mixing of the unburnt cold matter with the hot ashes is expected to occur and instabilities, like the Rayleigh-Taylor instability, develop. In this case the velocity of the burning front will be determined by the growth rate of these instabilities, being still subsonic but reaching soon values higher than those corresponding to the conductive velocity.

In 1D models, we rely on parametrization of this 3D turbulent mixing. The simple way is to assume a flame velocity equal to a few percent of the sound velocity [60]. The flame travels outward through lower and lower densities and would quench before reaching the surface. As quoted above, the observations require that all the WD is burnt; therefore, at some point, the flame velocity has to be increased as it happens in the delayed detonation models quoted before [39]. However, the physics of the transition to a detonation in unconfined environments, like in current Type Ia SN models, is not completely understood [61]. This transition is expected to depend on the properties of the structure, like the chemical gradient of fuel (*Zeldovich induction mechanism*) and, hence, on the progenitor.

Current 3D simulations are still far from a complete and successful explosion and even farther from being able to follow simultaneously all the relevant scales [62, 63, 64, 65, 66, 67]. As a consequence, successful explosions are still 1D simulations, being the assumed transition density (or the equivalent pre-expansion time) the main factor determining the amount of  $^{56}\text{Ni}$  produced and therefore, the maximum luminosity and LC shape.

### 5.3 Any observed hints from the companion?

A great observational effort is being done to identify the elusive companion of the exploding WD. First detections have been reported, although a clear answer has not yet been achieved. The spectra of 2002ic [68] show interaction with the circumstellar medium, hydrogen is abundant and an AGB star was identified as the companion. However both of the proposed scenarios (SD and DD) could explain this event [69] and, recently, it has been proposed that it was a core-collapse SN instead of a SNIa [70].

Interaction with the circumstellar medium is also detected in the case of 2006X [71] and a red giant has been proposed as the companion, favouring the SD scenario. However, note that H-lines are not seen in the spectra.

Another possibility is to look for the companion in nearby Type Ia SN-remnants, this has been done for the Tycho's supernova (SN 1572); Ruiz-Lapuente [72] claimed the identification of the donor star, Ihara [73] argue that this is not a viable object.

Finally, a very interesting case is the recurrent nova RS-Oph, for which the mass of the accreting WD is estimated to be very close to  $M_{Ch}$ ,  $\sim 1.38 M_{\odot}$ , and it is still increasing [74, 75]. Note that in this case there is H present in the system and in fact H is observed.

## 6 Influence of the progenitor of the exploding WD on SNIa properties

Let us now address the question of the influence of the WD progenitor on the final outcome. First, we will analyze the dependence of the final chemical composition of the WD on the initial mass of its progenitor, and then we will study how the light curves properties depend on this composition. Then, we will repeat the same analysis but considering the dependence of the WD progenitor properties on the metallicity. Note that given the *standard* model, a Chandrasekhar-mass CO WD, not much space is left for variations. In this set of models, the differences among the WDs are due to chemical composition differences: carbon and oxygen abundances and chemical gradients. This translates into differences in the  $^{56}\text{Ni}$  mass synthesized during the explosion and therefore, differences in the LCs.

We have considered the entire range of potential progenitors, with main sequence masses from 1.5 to  $7 M_{\odot}$  and metallicities between  $Z = 0$  and 0.02. The pre-supernova evolution, starting from the pre-main sequence phase, is simulated by means of a 1D hydrostatic code, the FRANEC code [76, 77, 78, 79]. Starting from the computed WD structures, we simulate the explosion, including detailed nucleosynthesis, and obtain the LCs [43, 80]. The single degenerate scenario has been assumed; the WD reaches the  $M_{Ch}$  by accretion of H from a non-degenerate companion.

### 6.1 Influence of the Initial Mass

We identify the initial mass of the progenitor (main sequence mass) of the exploding WD as a key parameter in the evolution; it modifies the average C/O ratio in the WD by up to 22% and thus, the kinetic energies and  $^{56}\text{Ni}$  masses [80].

We could distinguish two regions in the exploding structure: the inner one, the original WD (with masses between 0.55 and  $0.99 M_{\odot}$ , corresponding to initial masses between 1.5 and  $7 M_{\odot}$ ) and the external one, formed by the accreted matter. The C/O rate is determined by the previous He-burning phases. Initially carbon is produced via the  $3\alpha$  reactions and once sufficient  $^{12}\text{C}$  is synthesized, the  $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$  reaction becomes competitive with the  $3\alpha$ , and carbon is partially burned into oxygen. The amount of C/O is greater ( $\sim 1$ ) in the accreted matter because in these layers the  $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$  reaction does not have enough time to destroy  $^{12}\text{C}$  (unlike the situation during central convective He-burning). This higher C/O rate in the

previously accreted layers is critical for the final average C/O within the exploding structure.

The results are summarized in Table 1, where we report from column 1 to 6: (1) initial Z and Y, (2) initial (main sequence) mass, (3) mass of the CO core at the beginning of the thermal pulse phase, (4) C abundance (mass fraction) at the center, (5) averaged C/O rate within the final  $\sim 1.37 M_{\odot}$  CO white dwarf after accretion, and (6)  $^{56}\text{Ni}$  mass synthesized during the explosion.

More massive progenitors accrete less mass, as a consequence the final average C/O is smaller, less  $^{56}\text{Ni}$  is synthesized, the SNIa is less luminous, the LC-decline is faster and expansion velocities are smaller (up to 2000 km/s considering the whole range). Notice that the correlation between LC shape and expansion velocity could be used to further reduce the scatter in the calibration relation. The size of this effect on the maximum magnitude,  $\Delta M_{max} \sim 0.2$  mag, is critical for the reconstruction of the cosmological equation of state.

In Figure 1 the final C and O profiles (mass fraction) are shown for some selected models with different initial (main sequence) mass and Z.

	$M_{MS}(M_{\odot})$	$M_{CO}^{TP}(M_{\odot})$	$C_{cen}$	$C/O_{Mch}$	$^{56}\text{Ni} (M_{\odot})$
Z=0.02	1.5	0.55	0.21	0.75	0.589
Y=0.28	3.0	0.57	0.21	0.76	0.584
	5.0	0.87	0.29	0.72	0.561
	7.0	0.99	0.28	0.60	0.516
Z=10 <sup>-3</sup>	1.5	0.59	0.24	0.76	0.587
Y=0.23	3.0	0.77	0.26	0.74	0.567
Y=0.23	5.0	0.90	0.29	0.66	0.541
	6.0	0.98	0.29	0.60	0.522
Z=10 <sup>-10</sup>	5.0	0.89	0.32	0.70	0.549
Y=0.23	7.0	0.99	0.31	0.62	0.525

Table 1: Properties of the models.

## 6.2 Influence of the initial metallicity

As it is shown in Table 1, the mass of the CO core does not depend markedly on the initial metallicity of the progenitor; the only exception being the  $3 M_{\odot}$  that at Z=0.02 behaves as a low-mass star. As a consequence, the final chemical composition of the exploding WD is not sensitive to the initial Z [80]. For this reason, the  $^{56}\text{Ni}$  mass and the maximum luminosity depend mildly on the initial metallicity.

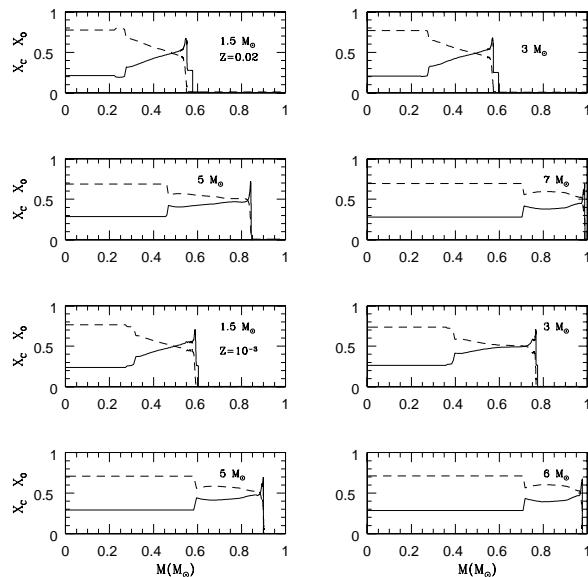


Figure 1: Final chemical Carbon (solid) and Oxygen (dotted) profiles in the central region of stars with initial (main sequence) masses between  $1.5$  and  $7 M_{\odot}$  and for  $Z = 0.02$  (upper pannels) and  $Z=0.001$  (lower pannels) (see Table 1).

However, the metallicity alters the isotopic composition of the outer layers of the ejecta and few Fe-lines contribute to the opacity. This effect is important at short wavelengths, changing the intrinsic colour index B-V by up to  $-0.06$  mag; moreover, it alters the fluxes in the U band and the UV [81]. The change in B-V is critical if extinction corrections are applied; the metallicity effect can systematically alter the estimates for the absolute brightness by up to  $0.2$  mag [80].

## 7 WD rotation and the properties at explosive C-ignition

In this case, we have assumed that two WDs (DD scenario) rotate at the orbital frequency due to the combined action of shear friction and tidal forces. Note that, once synchronization is attained, it will be maintained up to the merging epoch. The heating produced by the frictional stresses determines that two WDs heat up and

become more luminous during the GWR-driven shrinking [83]. Then, the two WDs merge so that the less massive component (the more expanded) completely disrupts, forming a thick disk from which matter flows to the companion. The real initial value of the accretion rate must be close to the Eddington limit. Fortunately, as we have shown, the initial value of this accretion rate does not affect the further evolution at all [82].

On the accreting WD, thermal energy is stored in the external layers while angular momentum is stored in the entire structure, which spins up. Due to the combined effect of these two processes, the accreting structure rapidly expands and accretion comes to a halt due to the onset of the Roche instability. Hence, the heat excess localized into the external layers is removed via thermal diffusion toward the inner zones. As a consequence the structure contracts and spins up. However, due to the redistribution of angular momentum, the angular velocity increases more slowly than the critical velocity and accretion is restored. It is as if there exists a gate which can be closed or open, depending on the actual physical conditions of the accreting WD. If such a process continued indefinitely, with no additional physics coming into play, the mean accretion rate would decrease with time and the self-tuned accretion mechanism would come to a definitive halt well before the Chandrasekhar mass is attained. However angular momentum is expected to be lost by gravitational wave radiation and the accretion process continues to occur in a self-regulated fashion until the Chandrasekhar mass limit for non-rotating structures has been already achieved or by-passed [82].

$M_{tot}$ ( $M_{\odot}$ )	$\rho_{ig}$ ( $\text{g/cm}^3$ )	$\omega_{rot}$ (rad/s)	$U_{bin}$ (foe) <sup>1</sup>	$M_{eje}$ ( $M_{\odot}$ )	$E_k$ (foe) <sup>2</sup>	$^{56}\text{Ni}$ ( $M_{\odot}$ )	$M_{bol}$
STD							
1.39	$2.90 \cdot 10^9$	0.0	-0.52	1.39	1.09	0.772	-19.48
$\tau_{bra}=10^4$ yr							
1.39	$2.17 \cdot 10^9$	1.07	-0.53	1.39	1.15	0.819	-19.58
1.42	$2.08 \cdot 10^9$	1.80	-0.56	1.42	1.17	0.847	-19.61
1.45	$2.09 \cdot 10^9$	2.27	-0.59	1.45	1.17	0.857	-19.61
1.48	$3.34 \cdot 10^9$	3.08	-0.66	1.30	1.04	0.817	-19.52
$\tau_{bra}=10^5$ yr							
1.39	$3.14 \cdot 10^9$	0.78	-0.53	1.36	1.08	0.774	-19.50
1.42	$2.84 \cdot 10^9$	1.92	-0.57	1.42	1.11	0.813	-19.56
1.45	$2.75 \cdot 10^9$	2.49	-0.61	1.44	1.12	0.837	-19.59
1.48	$3.01 \cdot 10^9$	3.02	-0.66	1.37	1.07	0.835	-19.55

<sup>1</sup> 1 foe= $10^{51}$  erg

Table 2: Properties of the rotating models.

Models with WD masses ranging between 1.4 and 1.5  $M_{\odot}$  have been analyzed. Differentially rotating WDs may be stable up to several solar masses but here rigid body rotation has been assumed so that the corresponding Chandrasekhar limit is  $\sim 1.5 M_{\odot}$ . As in the previous section, we assume the delayed detonation explosion mechanism, keeping constant the transition density for all our experiments, and compute the explosions and light curves [84, 85, 86]. This is the only reasonable choice to disentangle the effects of rotation on the observational properties of explosive events.

To slow down the WD a braking efficiency is assumed, physically motivated by the frictional viscosity due to the interaction of the WD with the accretion disk [82].

Our main results are summarized in Table 2: (1) total mass, (2) ignition density, (3) angular velocity at ignition-time, (4) total binding energy, (5) ejected mass, (6) kinetic energy (7) mass of  $^{56}\text{Ni}$  produced in the explosion and (8) bolometric magnitude at maximum time. Model *STD* is a reference model, without rotation, while  $\tau_{bra}$  refers to the adopted values for the time scale describing the braking efficiency.

At the time of explosive carbon ignition, the ignition density varies by 46%; the higher values,  $\sim 3 \cdot 10^9 \text{ g/cm}^3$ , are reached in the more massive models. This trend is also shown by the binding energies (with a variation of about 22%).

The amount of  $^{56}\text{Ni}$  produced depends on the binding energy and on the ignition density, and the two factors act in opposite directions. More massive models have greater binding energies and greater ignition densities; more energy has to be invested to unbind the WD, implying less final kinetic energy and a larger completely incinerated zone. However, electron captures are also favoured in the inner high density region, so that the production of  $^{56}\text{Ni}$  does not show a monotonic dependence with the ignition density, binding energy or total mass. The spread in  $^{56}\text{Ni}$  mass ( $\sim 0.08 M_{\odot}$ ) translates into a difference at maximum of 0.11 mag (see Figure 2).

## 8 Summary and conclusions

SNIa Hubble diagrams show the existence of dark energy and how the expansion rate of the Universe changes from deceleration to acceleration at  $z \sim 0.5$  [10]. SNIa absolute maximum magnitudes are obtained through a calibration relation that links them with the LC shape, the scatter in this relation is  $\leq 0.2$  mag.

- Observations have shown that the properties of the LCs and the rate of SNe Ia are related with the properties of their host galaxies; most luminous SNe Ia only happen in galaxies with ongoing star formation, in which their rate is also higher [35, 34]. This may indicate different stellar progenitors for Type Ia with evolutionary times that differ by 2-3 Gyr.
- Our theoretical studies have focused on the dependence of the SNIa properties on their progenitors. We have computed 1D numerical simulations, including

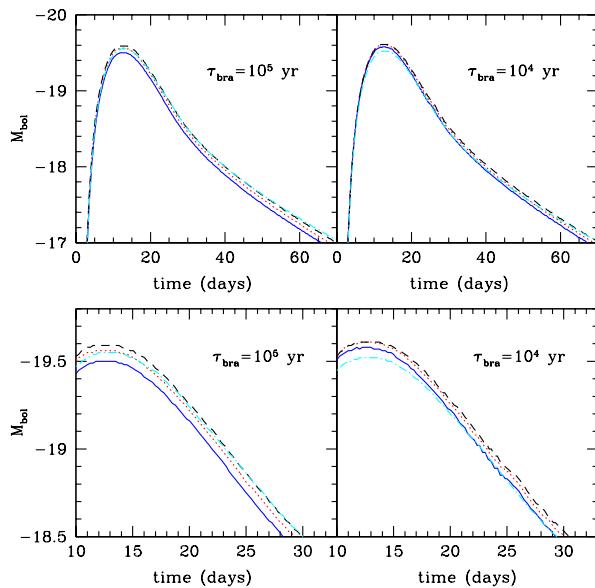


Figure 2: Bolometric light curves for the two set of models showed in Table 2. The spread in  $^{56}\text{Ni}$  mass translates into a difference at maximum of 0.11 mag.

stellar evolution, accretion onto the WD, explosions -assuming the delayed detonation mechanism-, detailed nucleosynthesis and light curves. We vary mass and chemical composition of the progenitor of the WD and, in the framework of the DD scenario, total mass of the system and rotational velocity. All these numerical experiments, in which explosion parameters (the crucial transition density) have been kept constant, led to variations at maximum  $\leq 0.2$  mag [80, 87].

- To characterize the EOS of dark energy using SNe Ia Hubble diagrams, the dispersion on the calibration relation should be reduced by a factor of 10. Otherwise we may have reached the precision limit of SNIa as distance indicators.

SNe Ia are the only astronomical observations that provide a direct evidence of the acceleration/deceleration of the Universe. It is worth to identify SNIa progenitors and explosion mechanism, to improve supernova cosmic distances and to analyze their reliability as precise probe of the dark energy nature.

*Acknowledgements:* I.D. warmly thanks Minia Manteiga and Ana Ulla for their enthusiasm and for their kind invitation to write this review. This work has been partially supported by the MCyT grant AYA2005-08013-C03-03, by the Spanish-Italian cooperation INFN-CICYT and by the Aspen Center for Physics (2007).

## References

- [1] Riess, A. et al. 1998, AJ 116, 1009
- [2] Schmidt B.P et al. 1998, ApJ 507, 460
- [3] Perlmutter et al. 1999, ApJ 517, 565
- [4] Krauss, L.M., Turner, M.S. 1995, General Relativity and Gravitation, 27, 1137
- [5] de Bernardis, P. et al. 2000 Nature 404, 955
- [6] Spergel, D.N. et al. 2003, ApJS 148, 175
- [7] Spergel, D.N. et al. 2007, ApJS 170, 377
- [8] Verde, L., Haiman, Z., Spergel, D.N. 2003, ApJ 581, 5
- [9] Schmidt B.P. 2005, ASP Conference Series, 342
- [10] Riess, A. et al. 2007, ApJ 659, 98
- [11] Kowal, C.T. 1968, AJ 73, 1021
- [12] Phillips, M.M. 1993, ApJL 413, L105
- [13] Phillips, M.M., Lira, P., Suntzeff, N.B., Schommer, R.A., Hamuy, M., Maza, J. 1999, AJ 118, 1766
- [14] Hamuy, M. et al. 1996, AJ 112, 2391
- [15] Perlmutter et al. 1997, ApJ 483, 565
- [16] Riess, A.G., Press, W.H., Kirshner, R.P. 1996, ApJ 473, 88
- [17] Norgaard-Nielson, H.U. et al. 1989, Nature 339, 523
- [18] Knop, R.A. et al. 2003, ApJ, 598, 102
- [19] Clocchiatti, A. et al. 2006, ApJ, 642, 1
- [20] Li, W. et al. 2003, PASP 115, 453L
- [21] Chornock, R. 2006 PASP 118, 722
- [22] Taubenberger, S. et al. 2008, MNRAS, in press (arXiv:0711.4548)
- [23] Howell, D. 2006, Nature 308, 283
- [24] Elias-Rosa, N. et al. 2006, MNRAS 369, 1880
- [25] Krisciunas, K. et al. 2004, AJ 128, 3034
- [26] Branch, D., Romanishin, W., Baron, E. 1996, ApJ 465, 73
- [27] Hamuy, M. et al. 1995, AJ 109, 1
- [28] Hamuy, M. et al. 2000, AJ 120, 1479
- [29] Ivanov, V.D., Hamuy, M., Pinto, P.A. 2000, ApJ 542, 588
- [30] Cappellaro E. et al. 1997, A&A 322, 421
- [31] Cappellaro E. 2003, *Supernovae and Gamma-Ray Burst*, Lec. Notes in Physics 598, 37
- [32] Wang, L., Hoefflich, P., Wheeler, J.C. 1997, ApJL 483, L29
- [33] Mannucci, F. et al. 2005, A&A 433, 807

- [34] Sullivan, M. et al. 2006, ApJ 648, 868
- [35] Mannucci, F., Della Valle, M., & Panagia, N. 2006, MNRAS 370, 773
- [36] Hoyle, W.A. Fowler 1960, ApJ 132, 565
- [37] Marion, G.H. et al. 2006, ApJ 645, 1392
- [38] Mazzali, P., Röpke, F.K., Benetti, S., Hillebrandt, W. 2007, Science 315, 825
- [39] Khokhlov, A. 1991, ApJ 245, 114
- [40] Khokhlov, A. 1995, ApJ 457, 695
- [41] Fisher, A., Branch, D., Höflich, P., Khokhlov, A. 1995, ApJL 447, 73
- [42] Höflich, P. 1995, ApJ 443, 89
- [43] Höflich, P., Khokhlov, A. 1996, ApJ 457, 500
- [44] Wheeler, J.C., Höflich, P., Harkness, R.P., Spyromilio, J. 1998, ApJ 496, 908
- [45] Lentz, E.J., Baron, E., Branch, D., Hauschildt, P.H., Nugent, P.E. 2000, ApJ 530, 966
- [46] Höflich, P., Gerardy, C., Fesen, R.A., Sakai, S. 2002, ApJ 568, 791
- [47] Höflich, P., Gerardy, C., Linder, E., et al. 2003, Lecture Notes in Physics (review), 635, 203
- [48] Becker S. A., Iben I. Jr. 1980, ApJ 237, 111B
- [49] Whelan, J., & Iben, I. Jr. 1973, ApJ 186, 1007
- [50] Hachisu. I. et al. 2007, ApJ 663, 1269
- [51] Piersanti, L., Cassisi, S., Iben, I.Jr., Tornambe, A. 2000, ApJ 535, 932
- [52] Piersanti, L., Gagliardi, S., Iben, I.Jr., Tornambe, A. 2003, ApJ 583, 885
- [53] Iben, I.Jr., Tutukov, A.V. 1984, ApJSS 54, 335
- [54] Webbink, R.F. 1984, ApJ 277, 355
- [55] Nomoto, K.I., & Iben, I.Jr. 1985, ApJ 297, 531
- [56] Saio, H., Nomoto, K. 1985, A&A 150, 215
- [57] Saio, H., & Nomoto, K.I. 1998, ApJ 500, 388
- [58] Napiwotzki, R. et al. 2005, ASPC 334, 375
- [59] Badenes, C., Hughes, J.P., Bravo, E., Langer, N. 2007, ApJ 662, 472
- [60] Domínguez, I., Höflich, P. 2000, ApJ 528, 854
- [61] Khokhlov, A.M., Oran, E.S., Wheeler, J.C. 1997, ApJ 478, 678
- [62] Gamezo, V.N., Khokhlov, A.M., Oran, E.S., Chtchelkanova, A.Y. Rosenberg, R. 2003, Science 299, 77
- [63] Gamezo, V.N., Khokhlov, A.M., Oran, E.S. 2005, ApJ 623, 337
- [64] Plewa, T., Calder, A.C., Lamb, D. 2004, ApJL 612 L37
- [65] García-Senz, D., Bravo, E. 2005, A&A 430, 585
- [66] Röpke, F. K., Hillebrandt, W. 2005, A&A 431, 635
- [67] Plewa, T. 2007, ApJL 657 942
- [68] Hamuy, M. et al. 2003, Nature 424, 651
- [69] Livio, M., Riess, A. G. 2003, ApJL 594, 93
- [70] Benetti, S., et al 2006, ApJ 653, L129
- [71] Patat, F. et al. 2007, Science 317, 924
- [72] Ruiz-Lapuente et al. 2004, Nature 431, 1069
- [73] Ihara, Y. et al. 2007, PASJ 59, 811

- [74] Hachisu, I. et al. 2006, ApJ 615, 141
- [75] Sokoloski, J.L., Luna, G.J.M., Mukai, K., Kenyon, S.J. 2006, Nature 442, 276
- [76] Chieffi, A., Limongi, M., Straniero, O. 1998 ApJ 502, 737
- [77] Domínguez, I., Chieffi, A., Limongi, M., Straniero, O. 1999, ApJ 524, 226
- [78] Straniero, O., Chieffi, A., Limongi, M. 1997, ApJ 490, 425
- [79] Chieffi, A., Domínguez, I., Limongi, M., Straniero, O. 2001, ApJ 554, 1159
- [80] Domínguez, I., Höflich, P., Straniero, O. 2001, ApJ 557, 279
- [81] Höflich, P., Wheeler, J.C., Thielemann, F.K. 1998, ApJ 495, 617
- [82] Piersanti, L., Gagliardi, S., Iben, I.Jr., Tornambe, A. 2003, ApJ 598, 1229
- [83] Iben, I.Jr., Tutukov, A.V., Fedorova, A. V. 1998, ApJ 503, 344
- [84] Bravo, E., Domínguez, I., Isern, J., Canal, R., Höflich, P., Labay, J. 1993, A&A 269, 187
- [85] Bravo, E., A. Tornambe, A., I. Domínguez, Isern, J. 1996, A&A 306, 811
- [86] Badenes, C., Bravo, E., Borkowski, K.J., Domínguez, I. 2003, ApJ 593, 358
- [87] Domínguez, I., Piersanti, L., Bravo, E., Tornambe, A., Straniero, O. Gagliardi, S. 2006, ApJ 644, 21



# GALAXIES AND THEIR ENVIRONMENTS: CONNECTIONS WITH NUCLEAR ACTIVITY

ISABEL MÁRQUEZ

*Instituto de Astrofísica de Andalucía-Consejo Superior de Investigaciones Científicas (CSIC), E-18008 Granada, SPAIN*

**Abstract:** One of the main issues concerning Nuclear Activity in galaxies (AGNs) is to understand the triggering mechanisms for the onset of non-thermal emission in their nuclei. Both the origin of the gas accreted onto the black hole and the physical mechanisms for the loss of angular momentum required for this funneling to be effective, have to be elucidated. In other words, the goal is to understand the needed conditions to switch on the AGN activity. But still many aspects of the investigation are a matter of debate. Among them, the rôle played by gravitational interactions and the relevance of the host galaxy need to be clarified. In this review, the different relationships between AGN activity, the morphological type of the host galaxy and the environment are discussed, in order to understand whether the AGN activity is more related to interacting effects or otherwise can be due to the secular evolution in the hosting galaxies.

**Keywords:** Active Galaxies – Structure – Environment.

## 1 Introduction

Most galaxies in the nearby universe belong to one of the Hubble morphological types, with two main groups of regular systems: elliptical galaxies, supported by the velocity dispersion of their stellar components, and disk galaxies, characterized by a bulge and a disk component (and very frequently also a bar), supported by circular rotation. There still exists a rather ill defined family: irregular dwarf galaxies. They are generally smaller, with low velocity rotation if any, and undersolar metallicities. Although they are very numerous in the Universe, dwarfs are irrelevant for the AGN nuclear activity population. AGNs are mostly related to the accreting processes taking place in the close environments of the black holes in the center of some galaxies [1], which is almost completely restricted to massive galaxies, i.e, to ellipticals and spirals.

The phenomenon of nuclear activity in galaxies have received much attention since its discovery, and already in the 80s the possible relationship of this kind of activity

with the host galaxy has been raised [2]. The connection seems to be related with the shape of the gravitational potential: earlier galaxies, with larger contributions from elliptical-like components, host AGN activity more frequently than galaxies with smaller such contributions (disk-like galaxies with small bulges). The properties of galaxies have long been known to depend on the environment in which they are located, with ellipticals mostly residing in rich clusters, and spirals mainly found in their outskirts (i.e. [3, 4]). Therefore, special care is needed when trying to disentangle internal from external drivers. The use of control samples matching both in morphology and environmental status will be crucial at this respect.

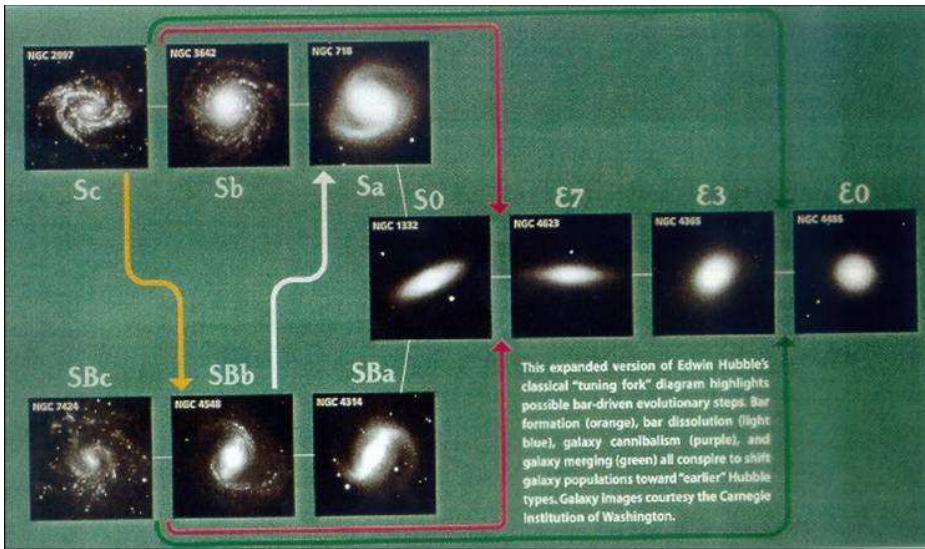


Figure 1: This figure has been taken from Martin & Friedli (1999) [5]. It illustrates how galaxies can change their morphological type by secular evolution, from later to earlier types. The formation of a bar due to disk instabilities provides the path from an Sc to an SBc. The rôle played by the bar in enlarging the bulge component and producing central mass accumulation, that destroys the bar itself, is responsible for the SBb to Sa transformation. The alternative path for transforming galaxies is given by the merger of two spirals, which ends up in an elliptical galaxy.

Galaxies have been proven to be dynamically evolving entities, whose properties can change with time both by secular processes and by the effects of interactions with neighbours or with the surroundings. Secular evolution has been suggested as a possible mechanism to make galaxies evolve from later to earlier types; this could

be the case specially within the group of spiral galaxies. The response of the disk to an initial, small external perturbation would produce gravitational instabilities in the disk giving rise to the formation of a bar, whose evolution gives rise to the transfer of material to the center; eventually, the bar itself will be destroyed due to the huge accumulation of gas. This effect is nicely illustrated in the figure by Martin & Friedli [5], reproduced here in Figure 1. Strong gravitational interaction can produce similar effects, but at more violent levels and, depending on the relative masses of the systems involved, giving rise to the formation of elliptical galaxies ([6, 7]).

AGN activity has to be sustained by some processes fuelling the nucleus and producing the plethora of properties observed at all wavelengths. Angular momentum loss has to operate in order to make the material to be funnelled to the center. The large scale processes driving material to the central regions have been explored as eventually related to those required much closer to the nuclear black hole. In this review, we analyse how AGNs are related to nature, i.e, the properties of the galaxy hosting an AGN, and how AGNs relate to nurture, i.e, with externally triggered modifications. All in all, we will keep in mind that the power of the AGN is another parameter to consider, as clearly illustrates the case of the strongest AGNs (quasars) corresponding to the most massive galaxies or the strongest interactions.

## 2 The rôle of the local/large scale environment

Many studies have dealt with the environments of AGNs, from few kiloparsecs to some hundred megaparsecs, with the aim of analysing whether these local or large scale environments are similar to those of inactive (those without AGNs, hereinafter non-active) galaxies. The analysis of the relation between AGN and environment has pointed to differing results. [9, 10, 11], among others, have concluded that no difference is found for the environment of active and non-active galaxies. Whereas other investigations suggest that the number of companions is higher for Seyfert galaxies than for non-Seyferts ([12, 13], with clear differences between Seyfert 1 and Seyfert 2 ([14, 15, 16, 17, 18]) and even depending on the power of the AGN ([19, 20]). The main limitations of these studies are related to the way the samples have been selected, how complete they are, how the different host properties are represented and, as it is also the case for other projects that need the definition of well defined control samples, how well these control samples (of non-active galaxies in this case) match with those of active galaxies. UV selected AGN samples could be biased against Seyfert 2 galaxies, since the obscuration expected to be more important in these objects would avoid or impede the detection of the nuclear source. This explanation was offered by [10] to explain the results by [14, 15, 16, 17, 18], that would contradict the predictions of the unified model (Antonucci 1993 [21]). They suggest that IR or X-ray selected samples would appear as more appropriated. In fact, very recently [22]

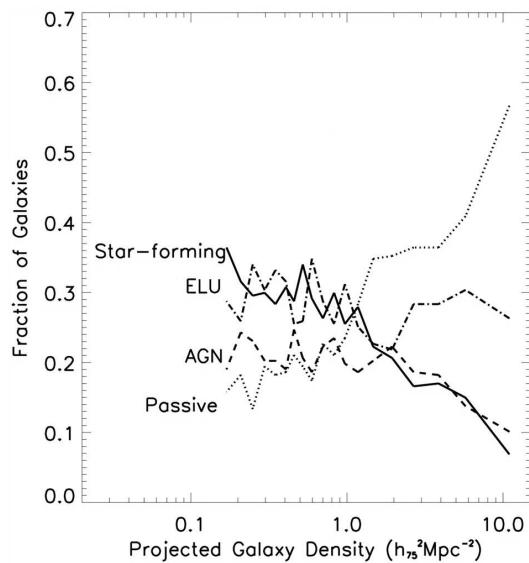


Figure 2: Fraction of the different galaxy classes (Star-Forming, Elusive, AGN and Passive), as a function of density. An increase in the fraction of passive galaxies with density is seen, together with a decrease in the fraction of star-forming galaxies with density. The fraction of galaxies possessing an AGN is statistically consistent with a constant over all local galaxy densities. Taken from Miller et al. (2003) [8].

have reported that the frequency of AGNs in X-ray selected clusters is much higher than previously found, explaining the paucity of AGN at optical wavelengths as due to obscuration.

The recent dramatic increase in the number of AGNs from several hundreds to several thousands, mainly from the SLOAN Digital Sky Survey (SDSS) has resulted in the possibility of approaching this study with statistical significance. The main result from the pioneering works with the first data release by [8] is that the fraction of AGN is found to be constant with projected galaxy density. In contrast, for higher densities the fraction of passive galaxies is enhanced and that for star forming galaxies decreases (Figure 2). This result is confirmed for different ranges in the relative velocity of the companion galaxies and for different redshift bins in the redshift range sampled, from  $z=0.078$  to  $z=0.05$ .

Kauffman et al. (2003) [23] analyse the properties of the host galaxies of a sample of 22,000 galaxies, through the stellar surface brightness ( $\mu^*$ ), the stellar mass ( $M_*$ ), the concentration index ( $C=R90/R50$ ) and  $D_{4000}$  as a proxy for the age of the stellar

population. They found that galaxies can be grouped into two families, corresponding to early types (ellipticals, lenticulars and Sa) characterized by  $3 \times 10^8 \text{ M}_\odot \text{ kpc}^{-2} < \mu^* < 3 \times 10^9 \text{ M}_\odot \text{ kpc}^{-2}$ ,  $C > 2.6$  and older stellar populations, and late types with  $\mu^* < 3 \times 10^8 \text{ M}_\odot \text{ kpc}^{-2}$ ,  $C < 2.6$  and younger stellar populations. The galaxies hosting an AGN, type 2 since broad line AGNs are not considered in this study, are almost exclusively massive galaxies, with the AGN fraction strongly declining for  $M^* < 10^{10} \text{ M}_\odot$ . They have similar sizes and stellar masses than normal early-types but show slightly different stellar ages than the parent general population. Depending on the power of the AGN source measured by the luminosity of the [OIII] emission line,  $L_{[\text{OIII}]}$ , they found that high power AGNs (those with  $\log(L_{[\text{OIII}]}) > 7.0$ ) show somewhat younger stellar populations (see Figure 3).

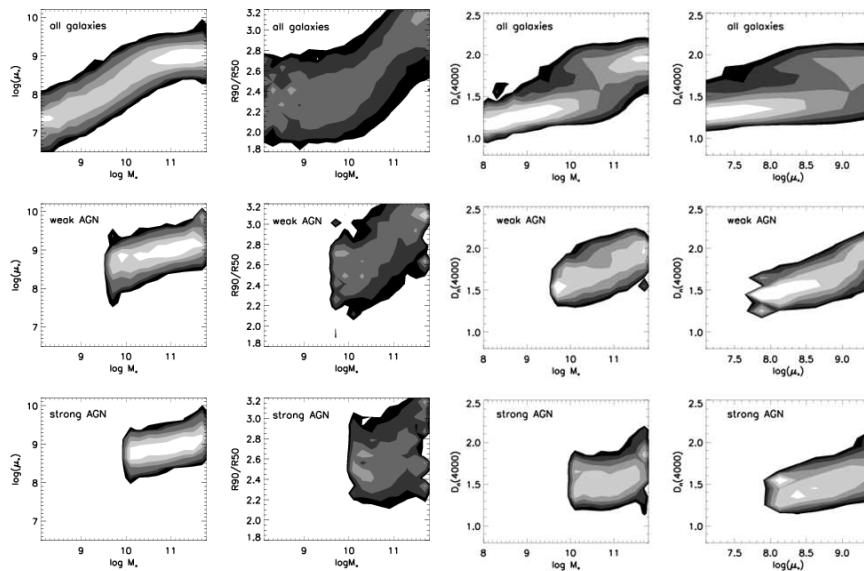


Figure 3: Conditional density distributions showing trends in the surface mass density  $\mu^*$  as a function of the stellar mass  $M^*$  for all galaxies (top), for weak AGN with  $\log(L_{[\text{OIII}]}) < 7.0$  (middle) and for strong AGN with  $\log(L_{[\text{OIII}]}) > 7.0$  (bottom). Taken from Kauffmann et al. (2003) [23], Figures 9 and 13.

With respect to the environment of AGNs, Kauffmann et al. (2004) [24] obtain a larger fraction of high power AGN in lower density environments, whereas at a fixed AGN power, the hosts are similar for any density (see Figure 4).

Sorrentino et al. (2006) [25] also consider the case of broad-line AGNs, analysing

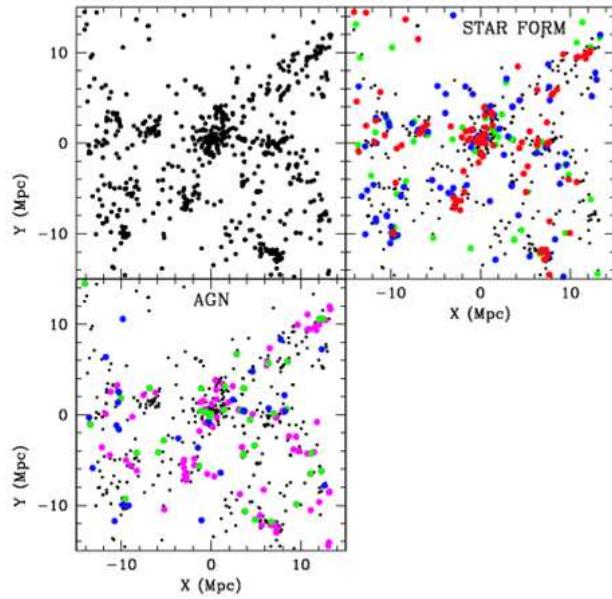


Figure 4: The distribution of all galaxies in a slice at  $z=0.05$ . Top right: galaxies with  $10^{10}M_{\odot} < M^* < 3 \times 10^{10} M_{\odot}$  are colour-coded according to their measured  $D_{4000}$  break strengths. Bottom: galaxies hosting AGNs are colour-coded according to  $L_{[OIII]}$ . Magenta is for  $\log(L_{[OIII]}) < 6.5$ , green is for  $6.5 < \log(L_{[OIII]}) < 7$ , and blue is for  $\log(L_{[OIII]}) > 7$ . Galaxies with no AGN are shown with small black dots. Taken from Kauffmann et al. (2004) [24], Figure 13.

a sample of 90,886 galaxies with  $0.05 < z < 0.095$  and  $M_r < -20$  from the SDSS, from which 2% are AGNs, 7% star-forming galaxies and 18% are considered passive galaxies. Based in the number density of neighbours in a volume with  $r_{max} < 100$  kpc, they conclude that Sy1 and Sy2 galaxies have similar large scale environments with a higher percentage of Sy2 appearing in close pairs (see Figure 5). Therefore, some of the conflicting previous results could be a consequence of not taking into account the necessity to distinguish between close and large scale environments when studying the eventual effects of the gravitational interactions as related to AGN activity.

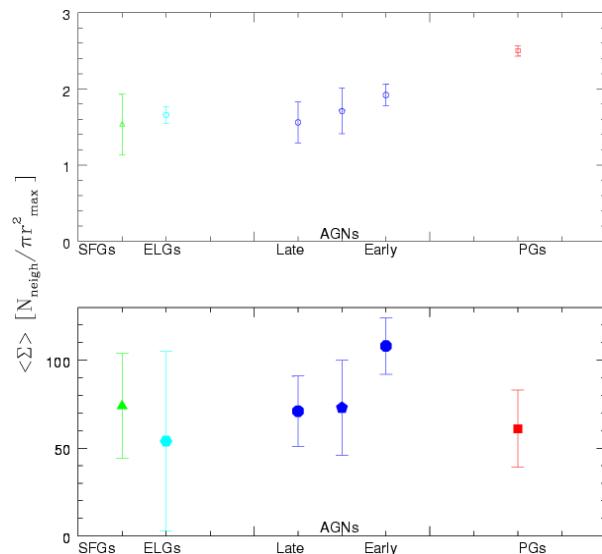


Figure 5: Mean surface density parameter for PGs (passive galaxies), SFGs (Star-forming galaxies) and AGN. Top: All systems with  $r_{\text{max}} < 100$  kpc. Bottom: Close systems ( $r_{\text{max}} < 100$  kpc). Taken from Sorrentino et al. (2006) [25], Figure 7.

### 3 AGNs and hosts

Previous to studies of huge samples in the last three years, the connection between the presence of AGN activity and the morphology of the host galaxy had been stressed from earlier studies ([2]). AGN hosts are more frequently found in early types, with a peak of the morphology distribution in Sb spirals ([2, 26, 27, 28, 29, 30, 31, 24]). Already in these works the necessity of explaining the internal mechanisms to be related to the onset of nuclear activity in galaxies was one of their first aims, and the presence of an asymmetric component of the gravitational potential was invoked as a main driver. Whereas in the case of interacting galaxies the departure from the symmetry is immediately provided by the tidal forces, the case of isolated galaxies deserved closer inspection, since the asymmetry should come from the host itself. But, at least in spirals, the presence of such internal asymmetric component is very frequent. Two thirds of spiral galaxies are barred ([32, 33, 34]) or even a higher proportion (up to 95%, [35]).

The importance of the presence of a bar in a spiral galaxy has been established already in the pioneering numerical simulations of barred galaxies, in which they un-

avoidedly appeared to stabilize the disk ([36, 37]). A bar is described as a component that rigidly rotates over a differentially rotating disk, what gives rise to different resonances usually associated to ring features (see Figure 6).

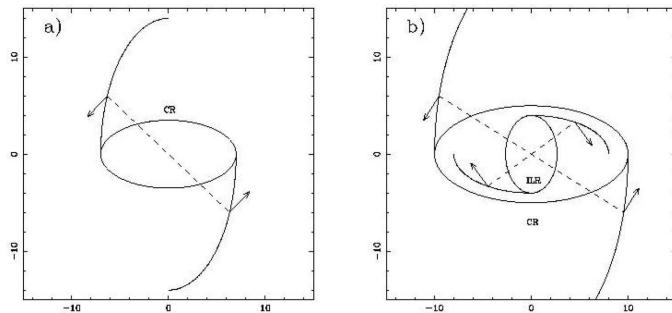


Figure 6: Illustration of Lindblad resonances in a barred spiral galaxy and the net flows they can produce: a) outflow between corotation (CR) and the Outer Lindblad Resonance (OLR); b) inflow between the CR and the Inner Lindblad Resonance (ILR). OLR and ILR are usually supposed to be traced by Outer and Inner rings, respectively.

Bars are easily formed in minor mergers ([38, 7]), and among their effects on the host galaxy, it is to note that they dynamically heat the disk, generate the development of spiral structure and produce net inflows to the central regions. Such inflows directly explain the observed flatness in metallicity gradient in barred galaxies, since the central, more metal rich, regions are contaminated by less metallic material coming from the outer disk. The infalling processes also relate to the various star forming features observed in barred galaxies, since they are expected to be different for old/young and strong/faint bars. Some of the important effects of bar secular evolution are discussed in [39], and summarised in Figure 7. The final stages of the infalling process, those of material accumulated unto the central region and bar destruction, have been explored, by means of numerical simulations, to possibly inducing the evolution in morphological types. Starting from a late type spiral that respond to a small perturbation of the disk, a bar is generated, which provokes the transport of some disk material to the center; this enhances the bulge component and destroys the bar itself, so that the galaxy finally ends up as an earlier unbarred galaxy (see Figure 1).



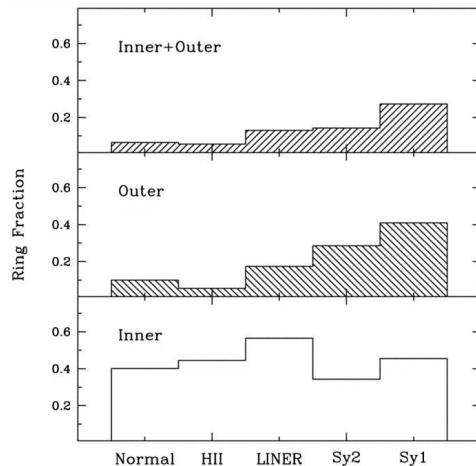


Figure 8: Fraction of ringed galaxies as a function of activity class. Lower panel: Inner rings. Middle panel: Outer rings. Upper panel: Galaxies with both inner and outer rings. Taken from Hunt & Malkan (1999) [40], Figure 4.

But still the question was whether there was a clearcut relation between the presence of a bar and the AGN activity. Whereas a consensus has not been reached at this respect (see [29, 41, 42, 43]), most works conclude that there is not an excess of bars among Seyfert galaxies ([27, 32, 28, 33]). Since bars evolve with time, some complications are expected in such a simple scenario, what could be related to the result by Hunt & Malkan (1999) [40] of a higher percentage of outer rings in Seyfert 1 galaxies (see Figure 8). The main concerns of all these works are related to the sample numbers, the sample selection procedures, and the way the control samples are defined. But in addition to all these eventual biases, the mechanism that is expected to drive the feeding material to the nuclear source has to operate down to the scales close enough to the nucleus, and large scale bars seem to be limited to produce such transport only till the region of the innermost resonance, at scales of about 1 kpc, where the material is trapped and no more inflow occurs. Several mechanisms have been proposed that could help in getting rid of the angular momentum at this point and drive the material closer to the center ([44, 45]). One of the first mechanism proposed as the required second step was that of nuclear bars, nested with respect to the large scale, primary bar ([46, 47, 48, 49, 50, 51]). Nevertheless, the results provided by the observations agree that no more nuclear bars are found in Seyfert galaxies ([52, 53, 54]) but on the contrary there seems to exist an excess of nuclear spirals ([55, 56, 57]) or nuclear disks ([58]) that can be stellar and decoupled from the

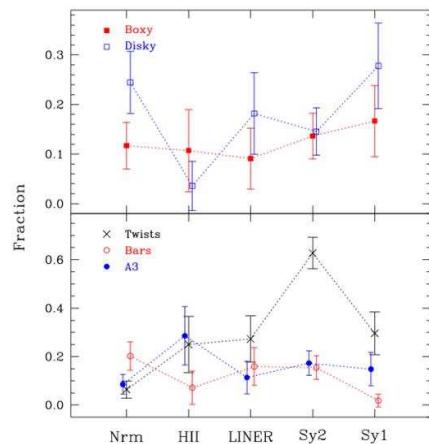


Figure 9: Fractions of nonaxisymmetric features as a function of activity types. A clear peak is found for twists in Seyfert 2 galaxies. Taken from Hunt & Malkan (2004) [60], Figure 8.

main galactic disk ([59]). Only when the presence of elongations is considered instead of that of a well defined bar, a slight excess (at  $2\sigma$  level) is found: central region of galaxies hosting HII, Seyfert 2, Seyfert 1 and LINER nuclei are progressively less asymmetric ([60]) a result that is interpreted in terms of an evolutionary scenario. It may be operating in such a way that these different types of nuclear activity would be connected in a temporary sequence (see Figure 9). Such evolutionary scheme seems to agree with that based on the results of the stellar population synthesis, which appear to be older when going from HII to Seyfert 2 to LINERs ([61, 62, 63, 64]).

Coming back to the possibility that interactions can provide the required asymmetry of the gravitational potential, and in an attempt to clarify the rôle played by the internal structure of the host galaxies, we started a project devoted to the characterization of ISOLATED Seyfert galaxies and its comparison with a matched control sample of ISOLATED spirals. The project DEGAS (Dynamics and nuclear Engine of Galaxies of Spiral type) was aimed at constructing a sample of nearby Seyfert galaxies from the Véron-Cetty & Véron (1993) [65] catalogue, with intermediate inclinations, with no reported belonging to any group or pair, with no companion at a projected distance of 600 kpc and with redshift difference smaller than 500 km/s, and with no projected companion in the DSS plates. The control sample was selected from the RC3 catalogue, imposing the same conditions for the absence of companions, and matching in redshift distribution, inclination and morphological type (including the

percentage of barred galaxies). The final samples amounted to 18 and 15 Seyfert and control galaxies, respectively. We analysed their NIR J and K images, better suited than optical images for tracing the gravitational potential, less contaminated by dust and with an expected smaller relative contribution of the AGN itself. The main results were that Seyfert and non-Seyfert hosts share similar bulges and disk properties (sizes, luminosities and surface brightnesses), but primary bars are also equivalent in both samples; secondary bars are found both in Seyferts (9 out of 12) and control (6 out of 10) galaxies. To characterize the gas kinematics, we obtained long slit spectroscopy along several position angles mainly in the wavelength range of  $H\alpha$  (see Figures 10 and 11). The kinematical properties of Seyfert and control galaxies appeared to be indistinguishable from those of early spiral types: same rotation curve shape, same position in the Tully-Fisher diagram, same disk metallicities, same kind of kinematical peculiarities in the central regions (Márquez et al. 2002 [66], 2004 [67]).

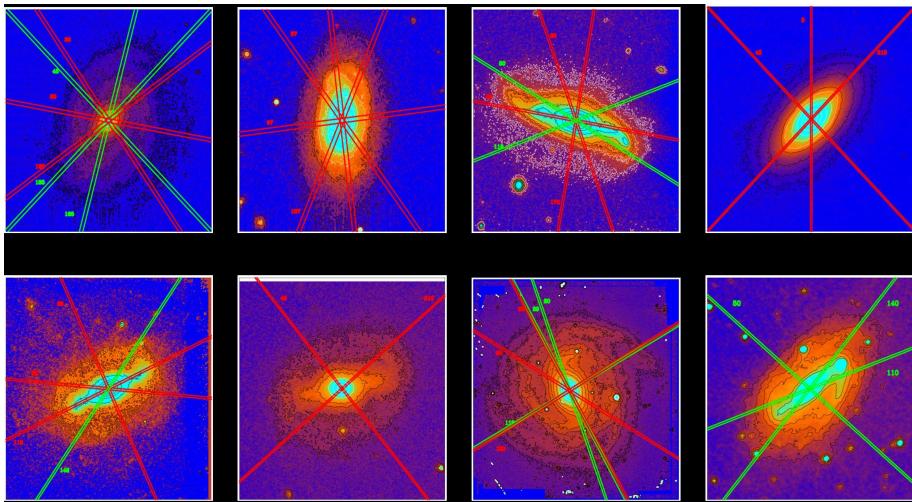


Figure 10: NIR images of some of the sample galaxies in the DEGAS project with the different observed position angles superimposed.

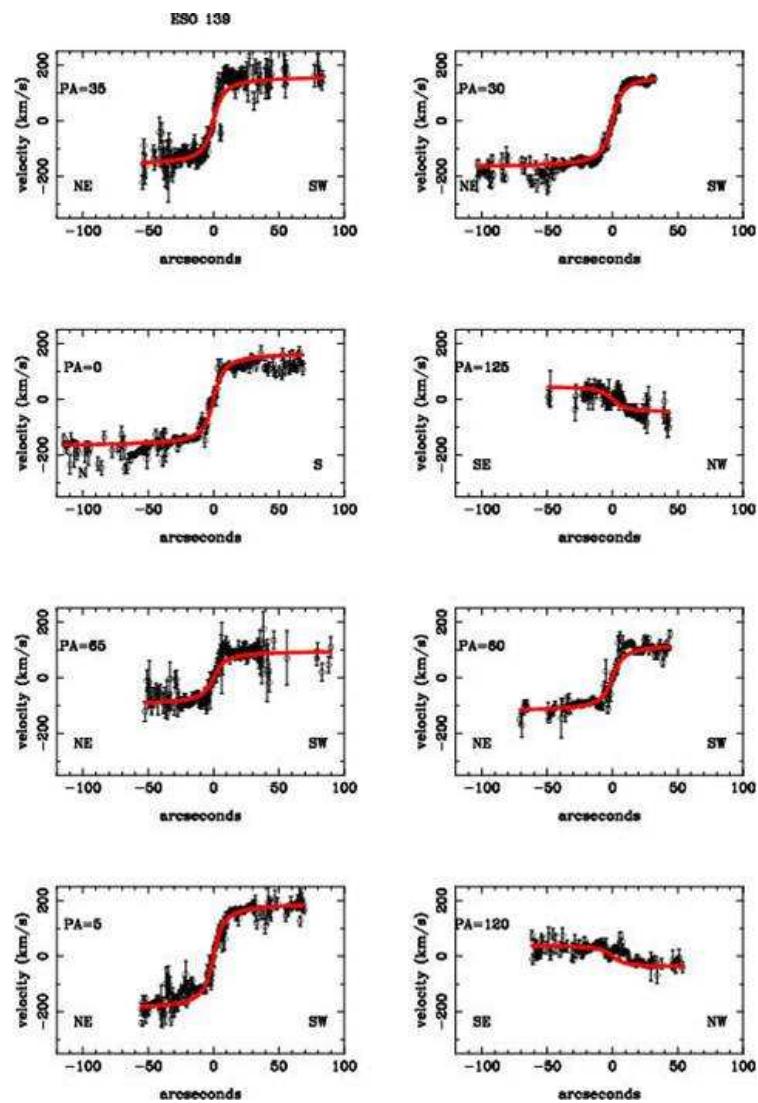


Figure 11: Velocity distributions obtained for one of our sample galaxies (ESO139-G12) along several position angles. A model for the disk rotation is plotted as the red line.

For deeper investigating how matter gets down into the nuclear region, a subsample was selected to study their morphological properties and the stellar and gaseous kinematics. We characterize the main properties of the host by using optical and/or NIR images from HST. The stellar kinematics was characterized with additional long slit spectroscopy along several position angles in the region of CaT, for obtaining stellar rotation curves and velocity dispersions and a stellar population tracer through the equivalent width of the CaT absorption lines,  $EW(\text{CaT})$ . Such study allowed us to detect the presence of a stellar velocity dispersion drop in the central 1-3 arcseconds in 5 galaxies, spatially coinciding with an increase in  $EW(\text{CaT})$ , that could hint the presence of young stars (red supergiants). Nine other galaxies in our sample had previously found to show such a drop. The analysis of the HST imaging of the total 14 galaxies showed that most of them showed a nuclear disk-like structure, spatially coinciding with the region where the velocity dispersion drop and the peak in  $EW(\text{CaT})$  occur (Márquez et al. 2003 [68], see an example in Figure 12). This result was interpreted in terms of the models by Wozniak et al. (2003) [69], that predict the formation of velocity dispersion drops once the gas coming from the outer, cooler disk, is driven to the center by the large-scale bar effects, giving rise to a decoupled nuclear disk. The gas velocity dispersion is hence smaller, so is that of the stars formed from it. Wozniak & Champavert (2006) [70] have recently updated the results from this modelling, and provided time scales for the whole process (less than 500 Myr for the formation of the drop, and lifetime dependent on the availability of fuel).

#### 4 A general picture for low-luminosity AGN hosts

A relatively recent consensus has been achieved on the presence of a black hole (BH) in the center of any massive galaxy ([71, 72, 73]), irrespective of whether it hosts an AGN or not. In addition, the properties of such BHs seem to be shared by both active and non active galaxies. Both types show the same relationship between the mass of the BH and the mass of the large scale spheroid hosting it ([74]). From the preceding discussion the properties of the host galaxies seem to be equivalent, both morphology and kinematics, at least the properties at scales of the order of the disk, bulge and/or bar components. The differences, if any, are expected to occur at scales much closer to the center, not still resolved by present day observations of the analysed samples. If even at much smaller scales no differences are found, an alternative explanation would be that the AGN activity can be switched on and off. This may be the case explored by the numerical simulations by Bournaud & Combes (2002) [75], that reproduce the evolution of galaxies, moving along and across the Hubble diagram, with any galaxy being able to become barred, or active, or both, and spend some time as an early type or late type. Nevertheless, a number of complications appear when trying to analyse in depth the different mechanisms related to the fuelling of low-luminosity

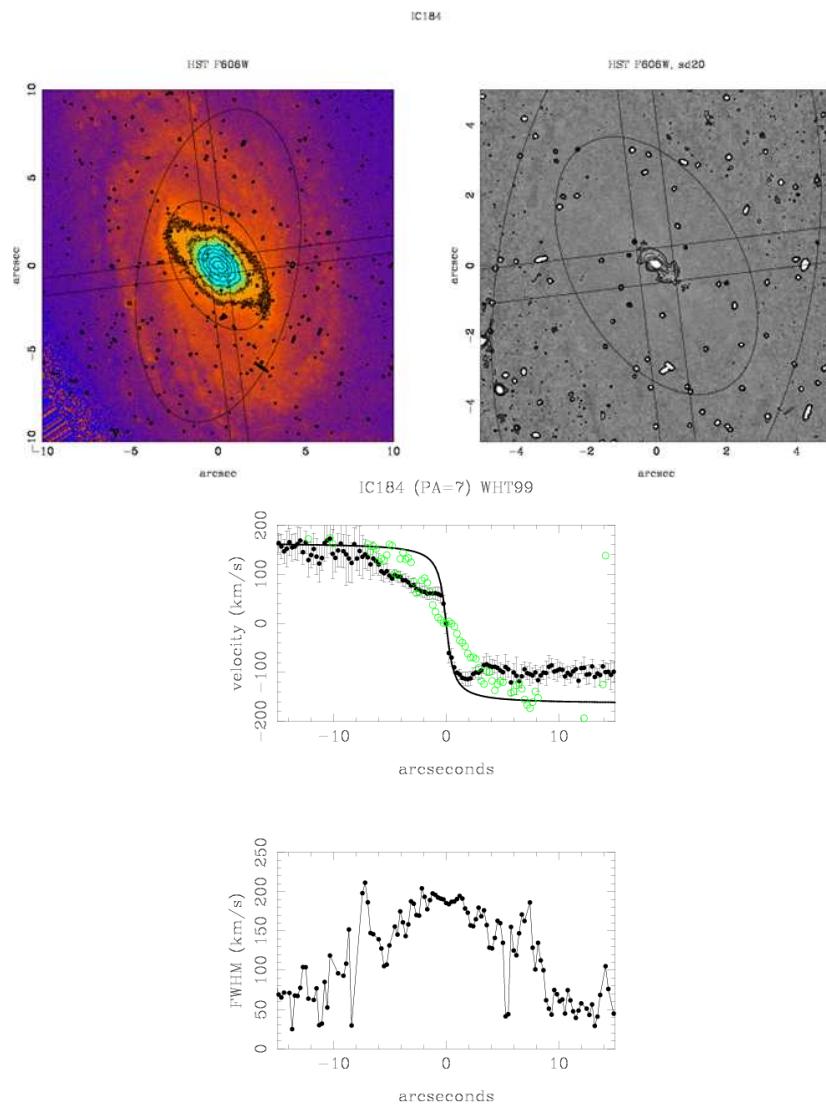


Figure 12: Top left: HST of IC184 in the F606W band with the two slits superimposed, the ellipses correspond (PA and ellipticities) to the two bars detected in the NIR. Top right: Sharp divided image of the center. Middle: Velocity curve of the gas (open, green circles) and stars (black circles) along PA=7. Bottom: FWHM of the stellar component along PA=7. Taken from Márquez et al. (2003) [68].

AGNs. Martini (2004) [76] proposed four reasons to explain why surveys have been unsuccessful up to now in resolving this question: (a) the current classifications for fuelling mechanisms are too broad, and additional refining is required, in particular for describing bars, since strong or faint bars are expected to produce different effects, (b) there are correlations between the fuelling mechanism and the fuelling rate, that are easier to identify for higher accretion rates (related to mergers), but deserves much closer inspection of the central parsec region at the lowest accretion rates (where other processes like dynamical friction on molecular clouds, stellar disruptions, many forms of turbulence, mass loss, etc, have an increasing relative importance), (c) multiple fuelling mechanisms may be operating, as it is the case for bars and interactions (but even when only isolated galaxies are considered, as it was the case for the DEGAS project, the results are not conclusive), (d) the two main time scales operating are the AGN lifetime and the fuelling time, so the time dependence is important and has to be taken into account before reaching any conclusions from the observations. The analysis therefore requires a broader description of the physical situation of the central regions, and the dynamical information is crucial at this respect.

The approach of the NUGA (Nuclei of Galaxies) consist on a very detailed analysis of the dynamical properties of a sample of nearby active galaxies using, in addition to the morphology, the 2D kinematics of the molecular gas with high spatial resolution. See two of their sample galaxies in Figures 13 and 14.

García-Burillo et al. (2005) [77] present such analysis for a small sample of galaxies, which allows them to derive an scenario for self-regulated activity in low-luminosity AGNs: an initial asymmetry in the disk would produce the formation of a bar, therefore a nuclear ring, where the infall of gas produced by the bar would accumulate gas; when massive enough, this circumnuclear ring would produce auto-destructive effects, weakening the bar, and allowing the viscosity in the ring to be responsible for additional inflow, the process finally ending in an axisymmetric configuration, again.

## 5 Future prospects

Both the detailed study of nearby galaxies and the statistical approach of massive surveys will be complementary in the study of the relationship between the structure of the galaxy, the environment and the power of the AGN it hosts. On one hand, detailed, high resolution studies of individual galaxies, with the requirement of providing dynamical clues of the different phases are still needed. For nearby galaxies, projects like NUGA are expected to produce fruitful results in the near future. High resolution imaging of a large number of distant galaxies has recently started to allow the exploration of how frequent single or even double bars are at  $z=0.1$  (see

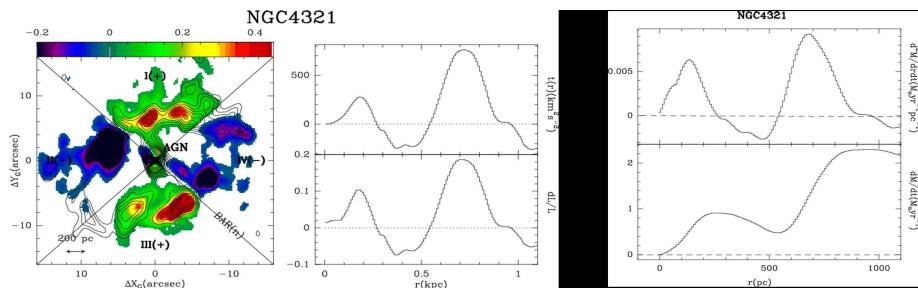


Figure 13: Left. 12CO(1-0) contours overlaid in the map of the effective angular momentum variation in the nucleus of NGC 4321. The derived torques change sign as expected from the action of the nuclear bar. Middle. The torque per unit mass averaged over azimuth -  $t(r)$  (top) - and the fraction of the angular momentum transferred from/to the gas in one rotation -  $dL/L$  (bottom)- are plotted. Torques are strong and positive for the bulk of the molecular gas, including the vicinity of the AGN. Torques are negative but comparatively weaker on intermediate scales ( $r=250-550$  pc) and in the outer disk  $r>900$  pc. Top right. Radial variation of the mass inflow(-) or outflow(+) rate of gas per unit radial length in the nucleus of NGC 4321 due to the action of stellar gravitational torques. Units are  $\text{yr}^{-1} \text{pc}^{-1}$ . Bottom right. Mass inflow/outflow rate integrated inside a certain radius  $r$  in  $\text{yr}^{-1}$ . The overall budget in NGC 4321 is clearly positive at all radii. From García-Burillo et al. (2005) [77]

[78], within the GOODS survey), what opens the possibility of studying evolutionary effects. On the other hand, a precious information remains to be extracted from existing massive surveys like SDSS, in terms of a much more detailed characterization of the morphology, not only with respect to the determination of precise morphological types instead of the rude parametrization by concentration indexes, but also on the presence and strengths of bars (see [79, 80] and FIGI project, [81]) and the characterization of the interaction state for AGN hosts and for comparable samples. Finally, focused numerical simulations with all the required ingredients, as those provided by Wozniak et al. for reproducing velocity dispersion drops, both for small (host) to large (environment) scales will help to understand the physical processes that give rise to the presence and onset of AGN activity in galaxies.

*Acknowledgements* IM gratefully acknowledges the SOC of this meeting for its kind invitation to contribute to it with this review, and J. Masegosa for her careful reading of this manuscript and interesting comments. Financial support from the Spanish Ministry of Science and Education, through AYA2003-0128, and the Junta

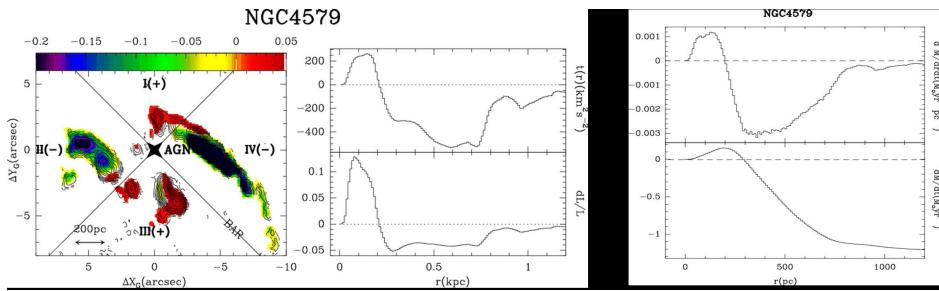


Figure 14: Same as Fig. 13 but for NGC 4579. The torques change sign as expected if the butterfly diagram, defined by the orientation of quadrants I-to-IV, can be attributed to the action of the large-scale bar. Torques are systematically strong and negative for the bulk of the molecular gas from  $r=200$  pc out to  $r=1200$  pc. In the vicinity of the AGN, however, torques become positive and AGN feeding is not presently favored. The overall mass inflow budget is clearly negative down to  $r=300$  pc due to the action of the large-scale bar. Inside this radius, stellar torques do not favour AGN feeding in this LINER/Seyfert.

de Andalucía is also acknowledged.

## References

- [1] Lynden-Bell, D. 1969, Nature 223, 690
- [2] Heckman, T.M. 1980, A&A 87, 152
- [3] Dressler, A. 1980, ApJ 236, 351
- [4] Balogh, M.L. et al. 1998, ApJ 504, 75
- [5] Martin, P. and Friedli, D. 1999, Sky & Telescope Vol. 97, No. 3, 32
- [6] Mihos, J.C., Richstone, D.O. and Bothun, G.D. 1992, ApJ 400, 153
- [7] Barnes, J.E. and Henquist, L. 1996, Ap. J. 471, 115
- [8] Miller, C.J. et al. 2003, ApJ 597, 142
- [9] Virani, S.N. et al. 2000, AJ 120, 1739
- [10] Schmitt, H.R. 2001, AJ 122, 2243
- [11] Schmitt, H.R. et al. 2001, ApJ 555, 663
- [12] Dahari, O. 1985a, AJ 90, 1772
- [13] Dahari, O. 1985b, ApJS 57, 643
- [14] Dultzin-Hacyan, D., Krongold, Y., Fuentes-Guridi, I. and Marziani, P. 1999, ApJ 513, 111
- [15] Krongold, Y. et al. 2001, AJ 121, 702
- [16] Krongold, Y. et al. 2003, Rev. Mex. A&A 39, 225

- [17] Koulouridis, E. et al. 2006a, ApJ 639, 37
- [18] Koulouridis, E. et al. 2006b, ApJ 651, 93
- [19] Hao, L. et al. 2005a, AJ 129, 1783
- [20] Hao, L. et al. 2005b, AJ 129, 1795
- [21] Antonucci, R. 1993, ARA&A 31, 473
- [22] Martini, P. et al. 2006, ApJ 644, 116.
- [23] Kauffmann, G. et al. 2003, MNRAS 346, 1055
- [24] Kauffmann, G. et al. 2004, MNRAS 353, 713
- [25] Sorrentino, G. et al. 2006, A&A 451, 809
- [26] Márquez, I. and Moles, M. 1994, AJ 108, 90
- [27] Moles, M., Márquez, I. and Pérez, E. 1995, ApJ 438, 604
- [28] Ho, L.C., Filippenko, A.V. and Sargent, W.L.W. 1997, ApJ 487, 568
- [29] Knapen, J.H., Shlosman, I. and Peletier, R.F. 2000, ApJ 529, 93
- [30] Maia, M.A.G. et al. 2003, AJ 126, 1750
- [31] Wake, D. et al. 2004, ApJ 610, 85
- [32] McLeod, K.K. and Rieke, G.H. 1995, ApJ 454, 77
- [33] Mulchaey, S.J. and Regan, M.W. 1997, ApJL 482, 135
- [34] Eskridge, P.B. and Frogel, J.A. 1998, AAS 193, 812
- [35] Grosbol, P. et al. 2004, A&A 423, 849
- [36] Hohl, F. 1971, ApJ 168, 343
- [37] Sellwood, J.A. 1981, A&A 99, 362
- [38] Hernquist, L. and Mihos, J.C. 1995, ApJ 448, 41
- [39] Jogee, S., Scoville, N. and Kenney, J.D.P. 2005, ApJ 630, 837
- [40] Hunt, L.K. and Malkan, M.A. 1999, ApJ 516, 660
- [41] Laine, S. et al. 2002, ApJ 567, 97
- [42] Laurikainen, E. et al. 2002, MNRAS 331, 880
- [43] Laurikainen, E. et al. 2004, ApJ 607, 103
- [44] Maciejewski, W. et al. 2001, MNRAS 323, 831
- [45] Jogee, S. 2006, in “Physics of AGN at all scales”, Lecture Notes in Physics Vol. 693, 143
- [46] Norman, C. and Silk, J. 1983, ApJ 266, 502
- [47] Shlosman, I. et al. 1989, Nature 338, 45
- [48] Wozniak, H. et al. 1995, A&AS 111, 115
- [49] Friedli, D. et al. 1996, A&AS 118, 461
- [50] Jungwiert, B. et al. 1997, A&AS 125, 479
- [51] Pfenniger, D. 2001, ASPC 197, 413
- [52] Regan, M.W. and Mulchaey, S.J. 1999, AJ 117, 2676
- [53] Márquez, I. et al. 1999, A&AS 140, 1
- [54] Márquez, I., Durret, F., Masegosa, J. et al. 2000, A&A 360, 431
- [55] Martini, P. and Pogge, R.W. 1999, AJ 118, 2646
- [56] Martini, P. et al. 2001, ApJ 562, 139
- [57] Shlosman, I. and Heller, C.H. 2002, ApJ 565, 921
- [58] Pogge, R.W. and Martini, P. 2002, ApJ 569, 624

- [59] Emsellem, E. et al. 2001, *A&A* 368, 52
- [60] Hunt, L.K. and Malkan, M.A. 2004, *ApJ* 616, 707
- [61] Boisson, C. et al. 2000, *A&A* 357, 850
- [62] Boisson, C. et al. 2002, *A&A* 396, 489
- [63] Boisson, C. et al. 2004, *A&A* 428, 373
- [64] Frémaux, J. et al. 2006, *A&A* 449, 109
- [65] Véron-Cetty, M. and Véron, P. 1993, “A catalogue of quasars and active nuclei”, ESO Scientific Report, Ic1993, 6th ed.
- [66] Márquez, I., Masegosa, J., Moles, M. et al. 2002, *A&A* 393, 389
- [67] Márquez, I. et al. 2004, *A&A* 416, 475
- [68] Márquez, I., Masegosa, J., Durret, F. et al. 2003, *A&A* 409, 459
- [69] Wozniak, H. et al. 2003, *A&A* 409, 469
- [70] Wozniak, H. and Champavert, N. 2006, *MNRAS* 369, 853
- [71] Magorrian, J. et al. 1998, *AJ* 115, 2285
- [72] Ferrarese, L. and Merrit, D. 2000, *ApJ* 539, L9
- [73] McLure, R.J. and Dunlop, S.J. 2002, *MNRAS* 331, 795
- [74] McLure, R.J. and Dunlop, S.J. 2004, *MNRAS* 352, 1390
- [75] Bournaud, F. and Combes, F. 2002, *A&A* 392, 83
- [76] Martini, P. 2004, in “IAU Symp. 222”, Cambridge Univ. Press, 235
- [77] García-Burillo, S. et al. 2005, *A&A* 441, 1001
- [78] Lisker, T. et al. 2006, *MNRAS* 370, 477
- [79] Ball, N. et al. 2004, *MNRAS* 348, 1038
- [80] Ball, N. et al. 2006, *MNRAS* 373, 845
- [81] Baillard, A. et al 2006, *ASP Conf. Series*, Vol. 351, p. 236

# MAGNETIC FIELDS IN GALAXIES

EDUARDO BATTANER<sup>1,2</sup>, ESTRELLA FLORIDO<sup>1,2</sup>

*1 Departamento de Física Teórica y del Cosmos, Universidad de Granada, SPAIN*

*2 Instituto Carlos I, Universidad de Granada*

ANA GUIJARRO<sup>1,3</sup>

*3 Centro Astronómico Hispano Alemán, Almería, SPAIN*

JOSE ALBERTO RUBIÑO-MARTÍN

*Instituto de Astrofísica de Canarias, La Laguna, Tenerife, SPAIN*

BEATRIZ RUIZ-GRANADOS<sup>1,2</sup>

ALMUDENA ZURITA<sup>1,2</sup>

**Abstract:** Magnetic fields not only provide a basic tool for observing galaxies but constitute an important clue to interpret many dynamic processes taking place in them, mainly in spiral and irregular galaxies. As in the Sun, magnetic fields become important where gravity becomes lower, i.e. in the outermost regions. Any peripheral feature should be explained taking into account magnetic fields, in particular warps, truncations of stellar systems, lopsidedness and so on. Of special cosmological interest is the role of magnetic fields in the rotation curve of spirals. All these questions are still insufficiently explored and the study of magnetic fields in the whole universe open wide areas of interpretation.

**Keywords:** Magnetic Fields – Galaxies

## 1 Introduction

The word magnetism has an Indo-European origin, coming from the root “magh” equivalent to “have power”. From the same root are the words “magic” and the spanish “desmayarse” (to faint). In non-scientific forums “magnetism” is still connected to the meaning of “magic” and even with “hypnotic” power to “fascinate”.

When magnetic fields lost their “magic power” to become a physical quantity? As we are aware the first scientific thought about magnetism being applied to a cosmic question was written in the logbook of a well known sailor around 1500:

“Fallo que de Septentrión en Austro, pasando cien leguas de las dichas islas [Azores], que luego en las agujas de marear, que fasta entonces noruesteaban, nordestean una cuarta de viento todo entero... Me puse a tener esto del mundo y fallé que no era redondo en la forma que escriben; salvo que es de la forma de una pera que sea toda

muy redonda, salvo all donde tiene el pezón, que allí tiene más alto, o como quien tiene una pelota muy redonda y en un lugar de ella fuese como una teta de mujer allí puesta...”. This was written after the third journey to America by Columbus.

There is a recent excellent book in which an exhaustive review of most aspects of galactic magnetism are deeply considered. This is the book “Cosmic Magnetic Fields” edited by Wielebinski and Beck in 2005. This book provides a so extensive and intensive up-to-date of galactic magnetism that it is not noteworthy to repeat hear the most established facts of galactic magnetism. The reader is also addressed to other interesting reviews [1, 2]. We will then concentrate in complementary topics concerning magnetic effects.

In this review we will concentrate in those topics concerning a) large scale magnetic phenomena and b) peripheric ones. This aim is equivalent to deal with controversial and even speculative topics. We will consider also magnetic fields in the pregalactic medium, as a possible clue to many features found today in  $z=0$  galaxies.

For other problems of galactic magnetic fields the reader is addressed to the cited book edited by Wielebinski and Beck. It is nevertheless interesting to briefly list some important topics at lower scale which will be here not considered, in which magnetic fields cannot be ignored or even are dominant, which are at present in phase of study, debate or promising development.

We will not consider the problem of AGN, radio lobes and optical and radio jets even if this is one of the chapters more insufficiently understood of galactic magnetism. Here, fields are not only the clue for explaining the ejection itself and the reacceleration of electrons but can provide an additional mechanism to feed the AGN [4]. The different dynamo mechanisms undoubtedly constitute one of the most important chapters, to explain how a regular field can arise as a consequence of turbulent motions, being a case where we “see” how chaos may produce order. Other small scale phenomena are as well important. Fields play a crucial role in HI and molecular clouds as well as in the HI diffuse medium, not only as a passive magnitude being amplified, ordered and disordered by them, and not only as a dynamic force, but also as a direct heat product, via reconnection [5] of magnetic field lines. This could be very important in other lower scale phenomena, such as in HII regions, bubbles and so on, in which the role of reconnection has just begun to be realized. Other phenomena as magnetic fields in bars and spiral arms, reversals and magnetic arms are not the objective here. Magnetic helicity can play an important role in our galaxy as well as in primordial magnetism. Finally, a considerable effort has been made in simulations in magnetized cosmic systems that will be not reviewed here [6, 3]. We will not consider the problem of the tools to observe and measure galactic fields. The synchrotron radiation constitutes the most important source of information about magnetic fields, but observing this radiation in some galaxy simply tells us that magnetic fields are there, but obtaining their strength and directions is a more difficult question. The synchrotron brightness not only depends on the magnetic field but also on the number

density and energy spectrum of the relativistic electrons. Therefore, some assumptions are needed, for example, equipartition of energies. The polarized synchrotron continuum provides clear information about the direction of the field vector, but only of its component perpendicular to the line-of sight. Faraday rotation also needs an additional observing tool to know the number density of the thermal electrons and we as well need a large number of extragalactic radio sources behind the observed galaxies. In our own Galaxy the number of extragalactic sources and galactic pulsars is high but the existence of a random component of the magnetic fields renders difficult the estimation of the large scale distribution of the regular component. (See Section 6).

## 2 Why are magnetic fields non-ignorable in any peripheric phenomenon.

The reason by which magnetic fields can be important in peripheric phenomena is double: a) They decrease slowly for increasing radii and b) Order of magnitude arguments show that they are not ignorable. The situation is therefore similar to that in the Sun and magnetic stars. Magnetic fields could compite with gravitation at large radii in normal spiral galaxies and at all radii in dwarf spiral galaxies.

There are evidences for large field strengths in galaxies, even at large radii:

- The field strength of a sample of 74 spiral galaxies is  $9 \mu\text{G}$  [7, 8]. Some galaxies, like M31, M83 and NGC6946 have a strength of the order of  $15 \mu\text{G}$  [8]. Starburst galaxies like M82 reach  $50 \mu\text{G}$  [9]. These are total fields but only the regular fields are able to produce magnetic forces at large-scale. Usually regular and random fields have similar orders, but at large radii the higher degree of polarization, and hence the field regularity, increases. At about 8 kpc the total field of NGC6946 is about  $10 \mu\text{G}$  [8] and the regular field must be slightly lower.
- The magnetic energy density,  $B^2/8\pi$  decreases very slowly with radius (Figure 1). For example, in NGC6946 (probably the best observed external galaxy [10]) the radial scale length of the magnetic energy density is very large, of the order of 16 kpc, compared with a radial scale of the density of this galaxy of only 3 kpc. The radial scale length of the regular field is probably larger than 16 kpc, as deduced from the higher degree of polarization at large radii.
- Faraday rotation of extragalactic sources behind M31 lead to magnetic strength similar to those found at lower radii [11].
- In our Galaxy the total field is  $6 \mu\text{G}$  at 10 kpc (about  $10 \mu\text{G}$  in the inner galaxy) [12, 13]. In Figure 2 we reproduce the radial profile [14] for our Galaxy. We

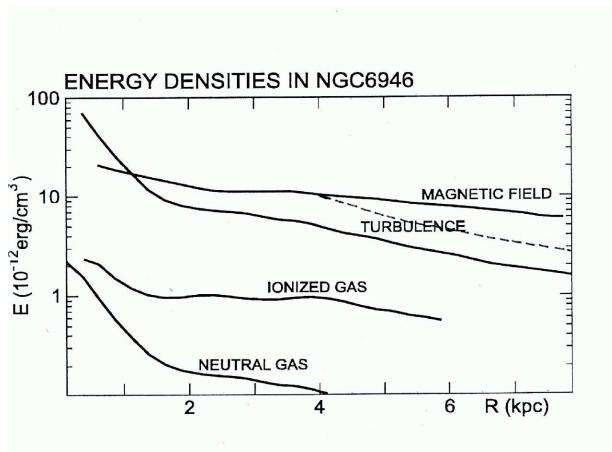


Figure 1: Energy densities in NGC6946 from [14]. The dotted line corresponds to a critical profile. The magnetic force is directed inwards.  $R$  is the galactocentric radius along the symmetry plane of the galaxy.

see that at 17 kpc the field strength of the total field is  $4 \mu\text{G}$ . Therefore, the regular field should be only slightly lower.

- The order of magnitude of the strength outside galaxies, in the intra-cluster medium, is similar. Faraday Rotation of sources placed behind the Coma cluster were detected [15] and deduced to be  $2 \mu\text{G}$  [16]; with much higher resolution it was obtained  $7\text{-}8 \mu\text{G}$  [17]. Statistically it was deduced a value of  $2\text{-}12 \mu\text{G}$  in a sample of 40 galaxies [18, 19, 20]. In addition,  $\mu\text{G}$  fields found everywhere, even in extragalactic regions near Coma could suggest an equipartition of magnetic and CMB energy densities [21]. In this case,  $B^2/8\pi = aT^4$ , being  $T$  the CMB temperature and  $a$  is the radiation pressure constant;  $a = 4\sigma/c$  where  $\sigma$  is the Stefan-Boltzmann constant. We should have for a widespread field,  $B \approx 3\mu\text{G}$ . That does not mean that a cosmological aligned field exists, what would be incompatible with the Cosmological Principle and in disagreement with the absence of a correlation between Rotation Measure in quasars and preferred directions of this relation. It has been estimated  $10^{-11}\text{G}$  as an upper limit for such homogeneous aligned cosmological field. However regular fields of the order of few microgauss are normal in clusters and may be present in large filaments and walls of the large scale structure [16]. Then, if the galactic magnetic field must connect with the intracluster field, and given the existence of an active turbulent magnetic diffusion, the regular magnetic field at large radii must be

as high as 1-5  $\mu\text{G}$  in normal spiral galaxies .

Field strengths this magnitude cannot be ignored. Or, in other words, if gravitational forces decrease as  $1/R^2$  (at large radii, and only due to the visible matter) and magnetic forces decrease much slower, the question is at what  $R$  the magnetic force becomes a substantial fraction of the gravitational one. Usually, the criterion for not neglecting a term is when it is at least one tenth of the dominant term. Hence

$$0.1 \frac{GM}{R^2} \rho \approx \frac{1}{8\pi} \frac{1}{R^2} \frac{d(R^2 B_\varphi^2)}{dR} \quad (1)$$

where  $M$  is the visible matter of the galaxy (seen at large radii as producing a central point potential),  $\rho$  is the density,  $B_\varphi$  is the azimuthal component of the magnetic field strength (the main component and the one that produces a radial force competing with gravity). For an order of magnitude estimate, and taking into account the above arguments, we assume  $B_\varphi$  independent of  $R$ , and consider  $\rho = \rho_o e^{-R/R_o}$ , then

$$R e^{R/R_o} = \frac{GM\rho_o}{B_\varphi^2} \equiv 4 \times 10^4 \frac{M\rho_o}{B_\varphi^2} \quad (2)$$

In the last term of this equation  $M$  is measured in  $10^{11} M_\odot$ ,  $\rho_o$  in  $10^{-23} \text{gcm}^{-3}$ ,  $B$  in  $\mu\text{G}$  and  $R$  in  $kpc$ . This transcendent equation would give us the radius at which magnetic fields cannot be ignored when compared with the visible mass of the galaxy.

We now take tentative values,  $B_\varphi \approx 5$ ,  $\rho_o \approx 1$ ,  $M \approx 1$ ,  $R_o \approx 3$  and obtain  $R \approx 14kpc$ . For some galaxies, the influence of magnetic fields can be very important, especially for dwarf spirals, if we take  $M = 0.1$ ,  $\rho_o = 0.1$ ,  $B_\varphi = 5$  (as dwarf spirals have even higher strengths [8, 22]) in which magnetic fields should be non-ignorable at all radii.

### 3 Magnetic fields and rotation curve

The techniques and difficulties for obtaining rotation curves have been detailly reviewed [23]. The possible influence of magnetic fields in the large scale dynamics of a spiral galaxy was early claimed [24, 25]. Later, a substantially different model was presented [26, 27]. In this and a subsequent model it was argued that magnetic fields alone, i.e. without the need of the hypothesis of galactic dark matter, were able to be responsible of the paradoxical large and flat rotation curve of spiral galaxies at the rim. Before reviewing the magnetic hypothesis of the rotation curve, it is salutary to prevent the reader that assuming negligible the contribution of galactic dark matter does not imply at all to assume negligible the amount of cosmological dark matter. On the other hand we cannot disregard the existence of galactic dark matter,

but the fact that existing models are able to explain this noticeable rotation curve, must prevent us that trying to explain it, ignoring magnetic fields, can be completely unrealistic. The magnetic hypothesis of the rotation curve is still controversial and a clear example of how much the astronomical community is prone to reject alternative scenarios, even when they are based on classical concepts.

A reconsideration of these magnetic models is now necessary as some of the properties of magnetic fields, such as strengths and radial gradients, which were assumed in these early models, have been recently confirmed by observations.

In a first model [26], it was claimed that fields of the order of  $6 \mu\text{G}$  in the outer parts could provide a centripetal force which together with the gravitational force of visible matter could balance the large centrifugal force. This “large” strength could produce a flaring higher than observed [28]. This important objection prompted us to develop a more detailed model [29]. In this two-dimension model that included horizontal and vertical motions and escape, only fields of  $\sim 1 \mu\text{G}$  at large radii, or slightly lower, were necessary. Today, with the development of measure techniques, it is not necessary to explain why an excessive flaring is not to be expected with fields of order  $5 \mu\text{G}$ , just because this is actually the observed strength. If the hydrostatic equation in the vertical direction leads to an excessive flaring, this just indicates that this equation was not properly applied. It was shown [29] that vertical motions and escape are basic ingredients in the vertical component of the motion equation.

The Virial theorem inform us about the net expansive effect of magnetic fields and however they can produce a centripetal force. No objection can arise from this fact as our model integrated the equation of motion while the Virial theorem is obtained from this equation under some restrictive assumptions. It was shown [29] that this potential incompatibility vanishes when the vertical direction is included. Magnetic fields should have a net expansive effect but they could have a centripetal action in the radial direction compensated with an escape in the vertical one. This second model predicted a vertical escape flux of the order of  $0.1 M_{\odot} \text{yr}^{-1}$ . Another recent work [30] argued against the hypothesis of magnetically driven rotation also revisiting the Virial theorem. Other arguments in favour or against have been discussed in [31] and will not be repeated here. Rotation curves under the interpretation of dark matter, MOND and magnetic field were analyzed and discussed [32]. These magnetically driven models [26, 29] were presented more than ten years ago. At that time several assumptions were made which have been clearly supported or confirmed by recent observations [31].

Magnetic forces arise from two different effects: the force due to the magnetic pressure gradients and the magnetic tension. The first one,  $\nabla(B^2/8\pi)$ , is usually centrifugal as  $B$  decreases for increasing  $R$ . The second one  $(1/4\pi)\vec{B} \cdot \nabla\vec{B}$  is, in real galaxies, centripetal. In a pure ionized gas ring, with an azimuthal field, it can be easily demonstrated that this magnetic tension produces an inward force. Then, the direction of the magnetic force arises from the competing action of gradients in

the magnetic energy density and the magnetic tension. The most important radial magnetic action, taking into account both effects can be written as a radial force

$$F_{radial} = -\frac{1}{8\pi\rho} \frac{d(R^2 B_\varphi^2)}{dR} \quad (3)$$

There are other forces but this represents the radial force produced by the azimuthal component of the field, which is the most important. We see that when  $B_\varphi \propto 1/R$ , there is no magnetic force. It is then expected that the  $B_\varphi$ -profile would asymptotically converge to a  $B_\varphi \propto 1/R$  profile for very large  $R$ , as the force should become vanishing when the density tends to zero. A profile of the type of  $B_\varphi \propto 1/R$ , or:

$$-\frac{1}{B_\varphi} \frac{dB_\varphi}{dR} = \frac{1}{R} \quad (4)$$

is called the critical profile. When  $B_\varphi$  decreases slowly with  $R$ , slowly than the critical profile, then  $d/dR(R^2/B_\varphi^2)$  becomes negative and the net force is centripetal. Such a profile is called sub-critical. On the other hand if  $B_\varphi$  decreases faster with  $R$ , the net magnetic force is centrifugal and the profile is called super-critical.

In the outer part of real galaxies we could identify three regions: A) An internal one where magnetic fields cannot compete with gravitation and therefore the shape of the  $B_\varphi$ -profile is unimportant, B) An intermediate zone where the  $B_\varphi$ -profile is subcritical and hence the net magnetic force is directed inwards, and C) An outermost region in which the  $B_\varphi$ -profile is critical, or better, asymptotically critical, in which the net magnetic force is null, corresponding to a density tending to zero.

There are at present only two galaxies for which we have data precise enough for this identification. These are NGC6946 and the Milky Way [14]. We reproduce the observational  $B_\varphi$ -profiles for both galaxies together with the critical slope (Figures 1 and 2 from [31]). In NGC6946 the profile is clearly subcritical for  $R \geq 3kpc$  indicating a net centripetal force. In the Milky Way we see that the slope is critical for  $R \geq 10kpc$ , and subcritical for  $7 \leq R \leq 10kpc$ . For  $R < 7kpc$ , clearly, magnetic forces should be negligible.

These two plots are in noticeable agreement with the theoretical expectations from the magnetic model and encourage theoreticians and observers to appreciate how magnetic fields should be included in the study of rotation curves, mainly when we see that standard  $\Lambda$ CDM models do not satisfactorily explain them ([33, 34, 35]). Some elliptical galaxies permit the obtention of rotation curves too. As they are gas-poor magnetic fields cannot be important. Rotation curves in these galaxies seem to be keplerian [36, 37].

Other types of galaxies should be investigated for large magnetic dynamical effects:

- Some galaxies exhibit an outer ring (both ellipticals and spirals). An isolated ring should be much more affected by magnetic fields as the force arising from

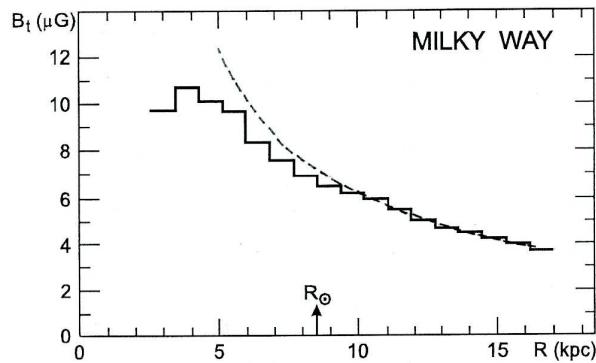


Figure 2: Total magnetic field strength in the Milky Way. A critical profile (dotted line) follows a region with magnetic force directed inwards. From [14] and [31].

the gradient of the magnetic energy density is absent. Therefore, the inwards magnetic force due to the magnetic force should be higher and easier to be identified. We include here polar ring galaxies, whatever was the origin of the ring.

- Dwarf irregular galaxies can be considered extreme late-type spirals. As they are small, according to the Tully-Fisher relation [38] the rotation velocity is low. With a low rotation velocity, dynamo models predict low magnetic fields strengths. However, observations show that these strengths are at least as high as in normal spirals, as above mentioned. Not only the peripheric regions could be influenced by magnetic fields, but the whole irregular galaxy. It should be stressed that magnetic fields also induce large vertical motions and that these galaxies exhibit especially vertical outflows as well as gas and field ejections. A good example is the galaxy DDO 154 [39]. Slowly rotating galaxies of the Local Group reveal strong total magnetic fields of more than  $10 \mu\text{G}$  (in NGC4449 and in IC10 ) [40].

Other works also present analyses favoring the existence of dynamically non-negligible magnetic fields [36, 41]. For a more complete review the reader is addressed to [32, 31] where arguments in favor and against are analyzed in more detail.

Another complementary problem is the identification of the mechanism producing these fields. The first approach would be how to obtain critical profiles. The possibility of a turbulent magnetic diffusion of the  $z$ -component of the magnetic field, amplified and converted into azimuthal field in the central disk by differential rota-

tion to produce critical  $1/R$ -profiles was considered [42, 32, 43, 44]. However the azimuthal component should vanish at  $z=0$  in the galactic plane contrary to the observations [44, 45]. See, however [11]. Other dynamo-like mechanisms are able to produce critical profiles [32].

## 4 Truncations of stellar disks.

Truncations of stellar disks were discovered and studied by van der Kruit and Searle [46, 47, 48, 49]. Other more recent observations have been reported in [50, 51] and many others. We understand that “truncations” should be “complete”. With the word “complete” we understand that beyond a certain radius,  $R_{max}$ , the density of the stellar system is null (within observational limits). This was the first description proposed by van der Kruit and Searle and this interpretation was followed and adopted by most astronomers [52, 53]. Clearly, truncations cannot be completely sharp, only relatively sharp [54]. Therefore, we can speak of a truncation curve,  $T(R)$ , defined as the difference between the observed light profile and the extrapolated exponential of the inner untruncated disk. Clearly, when we are observing edge-on galaxies what we observe is the surface brightness but the deprojected surface brightness should be proportional to the density of stars. We understand that truncations are peripheric, complete and not completely sharp.

From the observational point of view, several recent papers [55, 56, 57, 58, 59], have found that the “truncation” is not a good description of the phenomenon. They find that the correct description is a double exponential profile with a break in between. Sometimes, the second exponential is steeper, sometimes it is shallower and sometimes the profile is unbroken. This simple scheme proposed in [55] has been enriched in more recent papers where the authors divide these three types into a series of subtypes, often introducing a third exponential which can be either shallower or steeper than the second, rendering the initial classification more complex. When the second exponential is shallower they also use the word “untitruncation”. Most of their analysis is based in face-on spirals and in the optical, but edge-on galaxies also in the optical have also been studied [59].

A description more in agreement with the early interpretation of van der Kruit and Searle and Binney and Tremaine has been observed [60, 61, 62]. Following these authors there is a real non-sharp complete truncation being the truncation curve given by a fitting formula proportional to  $(R_{max} - R)^{-\alpha}$  where  $\alpha$  is close to unity. Figure 3 shows a typical truncation profile and in Figure 4 we plot the truncation curve. As these authors work in the NIR and the formers in the optical, the first explanation would be that both wavelengths, tracing different populations, are intrinsically different. However, the galaxy NGC6504 was observed in the optical and in the NIR [63] with the best depth available and it was found that the optical profile is as well

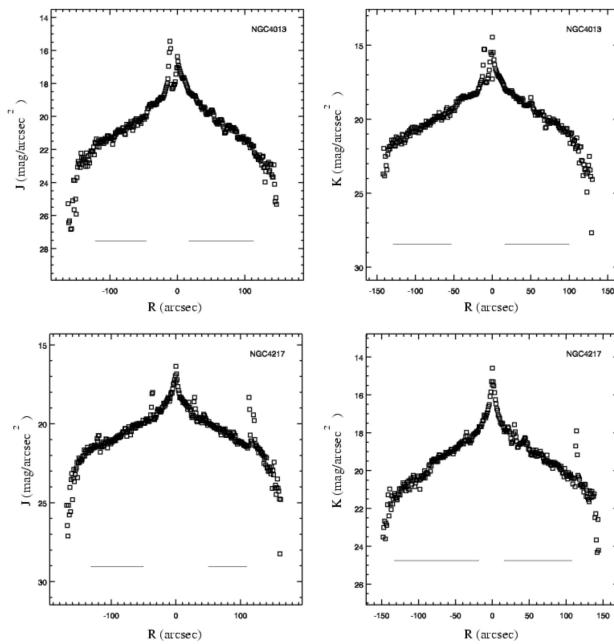


Figure 3: Truncation profile in the spiral galaxy NGC4013 and NGC4217 ([60]).

really truncated. They have reached 28.5 magnitudes per arcsec square with no trace of stars beyond  $R_{max} = 82$  arcsec. Profiles in the optical and in the NIR have a slope continuously steepening without evidence of a double exponential. (See Figure 5).

The NIR profiles also confirm the existence of antitruncations such as that of NGC2654 [62] as shown in Figure 6.

We are here, however, more interested in the physical process responsible of truncations. This section is included in this review because magnetism is one of the proposed mechanisms [64]. The scenario is simple and connected with that of magnetically driven rotation (see last section). Suppose gas moving in circular orbits with gravity plus magnetic fields in balance with the centrifugal force. Suddenly (in a very short time compared with typical dynamic times) some stars are born from this rotating gas. Magnetic fields can no longer act on the stars. Therefore new born

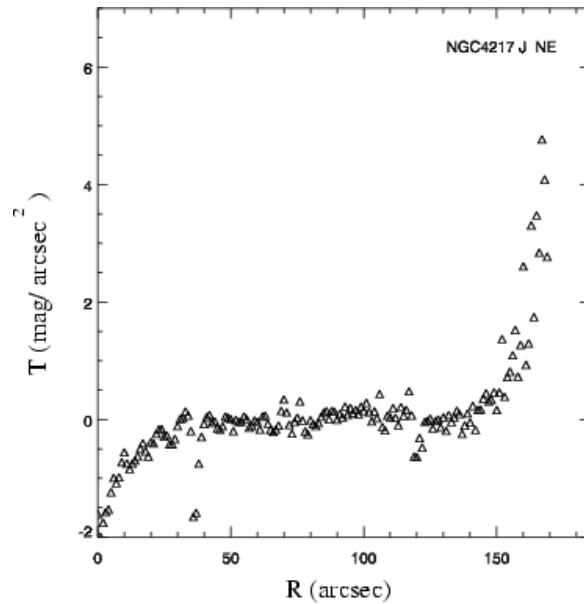


Figure 4: Truncation curve in the spiral galaxy NGC4217 ([60]).

stars are formed out of equilibrium of forces. Then, some stars migrate to larger radii orbits (producing antitruncations) and others escape from the galaxy in the radial direction (producing truncations).

This hypothesis has the “a priori” advantage that, qualitatively at least, is able to explain both antitruncations and truncations. For the case of truncations, we deduced [64] by means of our simpler model that

$$R_{max} = \frac{2GM}{\theta_o^2} \quad (5)$$

being  $M$  the visible mass of the galaxy and  $\theta_o$  the asymptotic rotation velocity ( $\theta_o = \theta(R \rightarrow \infty)$ ). Taking into account the Tully-Fisher relation  $L \propto \theta_o^x$ , where  $x$  is typically 3.5, we obtain either  $R_{max} \propto \theta_o^{1.5}$  or  $R_{max} \propto L^{0.7}$ . The relation between  $R_{max}$  and  $\theta_o^{1.5}$  is fully confirmed by observations. (See Figure 7 and [54] for edge-on galaxies). We also see that truncations should be more difficult to be detected in large galaxies, as  $R_{max}$  is higher. This is also confirmed by the observations.

There are other non-magnetic models for the explanation of truncations. If we assume that the initial gaseous disk was exponential without neither truncations nor

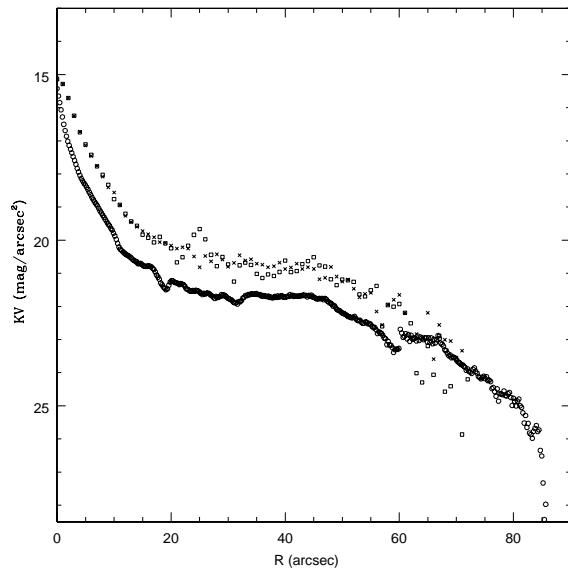


Figure 5: Truncation curves in the spiral galaxy NGC6504 in the optical and NIR ([63]).

breaks, and if the density of stars drop beyond an  $R_{max}$ , either, for some reason, the gas beyond  $R_{max}$  has no ability to form stars, or the stars once formed flow away. Below a density threshold, stars could be no longer formed out of gas. The threshold hypothesis [65, 66] is of the former type. The excess of angular momentum induced by a large differential rotation beyond  $R_{max}$ , could cause the inability to form stars too [67]. This explanation would also be of the former type. Clearly, magnetically driven truncations belong to the second type of explanations.

There is an interesting argument [68] against the scenarios inhibiting star formation. A sudden drop in the rotation curve takes place at  $R_{max}$ . This drop was also found in NGC5055 [69].

This drop indicates that there is a rather sharp decrease in the total mass (gas plus stars) not only of stars. Another argument favoring models which do involve large scale dynamics, is that the radial decrease of gas amount is due to a radial decrease in the number of molecular clouds, but the density in these outer clouds is not different when compared with the inner clouds.

Another interesting model, also involving large scale motions, has been proposed

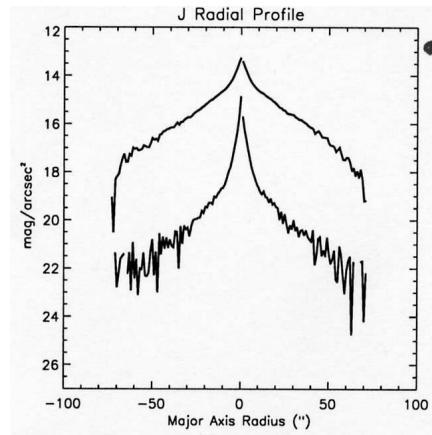


Figure 6: Antitruncation in the NIR (J band) in the galaxy NGC684.

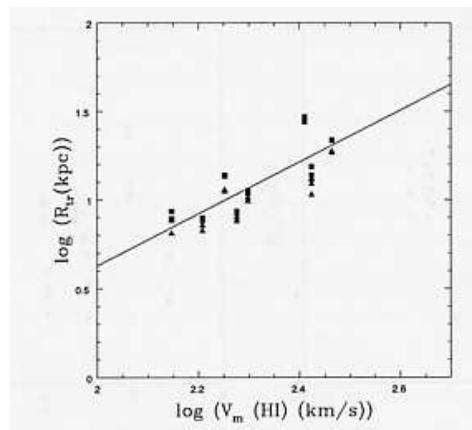


Figure 7: Relation between the truncation radius and the asymptotic velocity.

[68, 54], in which the disk was formed in two steps. In the first step the inner disk was formed with a size  $R_{max}$ . This finite size  $R_{max}$  was established by the maximum specific angular momentum. The extragalactic gas falling later, in the second step, produced the outer region.

With respect this model defended in [54] with the first settled truncated rigid disc and a warped external disc in a more recent step, relating  $R_{max}$  and the radius

at which the warp begins, we should have two comments: First, optical warps exist and even most galaxies have warped stellar systems [70, 71, 72]. Second, very early  $z=1$  galaxies show optical warps and much larger than the present one [98].

An important observational fact is that  $R_{max}$  increases with the band wavelength used (see for instance Figure 7). This is another reason to use NIR to detect truncations. This is also a constraint that models must explain. It is interesting to note that breaks are also observed in large- $z$  galaxies [74, 58].

## 5 Warps.

It is well known that most spiral galaxies have a warped plane when observed at 21 cm ([75, 76, 77] and others). The warps are also observed in the optical ([70, 72, 60, 71, 79]). The frequency of warped disk in spiral galaxies is very large. In contrast, no lenticular galaxy has been observed to be warped [72]. All the spiral galaxies with very extended HI-disk are warped. These facts suggest that the mechanism producing the bending acts directly on the gas and that the stellar warp is the consequence of being stars born in a warped disk. Therefore, we have searched for a magnetic scenario being in this case the extragalactic magnetic fields responsible of the warps.

There are other hypothesis (see [80, 81, 68], for instance). Tidal interaction [82] may explain some warps but not their high frequency: Near all spiral galaxies are warped. Sparke and Casertano [95] showed that some discrete modes can survive, but they did not take into account the reaction of the halo, what would destroy the warp in few orbit periods. Binney [80] considered the interesting possibility that the permanent accretion of material into the DM halo would produce a continuous redistribution of its angular momentum and the warps would be created by the reaction of the disk to this redistribution.

It seems that warping could be a natural response of the outer disk to a series of stimulations and that different mechanisms could be responsible in different galaxies. However the observation of large edge-on spiral galaxies could contribute to determine the dominant scenario. The intergalactic medium could play an important role [83] by interaction with extragalactic clouds [84, 85], by the flowing of galaxies in an intra-cluster medium [86] or by extragalactic magnetic fields [87].

The magnetic model of warps is an interesting possibility. It is known that a warp will disappear in a time of about 2 Gyr [80], or even shorter in the magnetic case. However, in the magnetic model of warps, the field configuration is not an initial condition but extragalactic fields are acting permanently. Any cause of warps must act permanently.

The intergalactic magnetic field must penetrate in the outer disk by some mechanism at work in the external region. Probably the interchange of matter and fields between disk and intracluster medium is very active as vertical out and inflows are

very common, transporting these flows the frozen-in field lines. This interchange can be very active mainly in the outer disk where the low surface density produces low gravitation enabling vertical motions. These flows could be considered as turbulent (even if they are not properly turbulent) and the mechanism of field transport could then be treated as a turbulent magnetic diffusion. The coefficient of turbulent magnetic diffusion,  $\beta = (B'/B)v'l$ , should not be constant but highly radial dependent.  $B'$  and  $B$  are the fluctuating and mean magnetic field strengths,  $v'$  a characteristic velocity of the fluctuations and  $l$  a typical eddy size (e.g. [88]). We take  $B'/B \sim 1/3$ , as the random component is slightly lower than the regular one [8] and  $l \approx 1$  kpc is the size of the larger eddies, that can be identified with the thickness of the disk. It is difficult to assess values for  $v'$  but it must change very much with radius. Taking  $v' \sim 1$  km/s in the inner disk we obtain  $\beta$  (inner) =  $10^{26}$  cm<sup>2</sup>s<sup>-1</sup>, which is even larger than currently adopted values [89, 90, 91].

The turbulent process producing the turbulent diffusion of magnetic fields is usually assumed to consist in bubbles arising massive type II supernovae which reach typical velocities of about 100 km/s and heights around 1 kpc from the plane, usually raining back [92]. Observation of bubbles, holes or chimneys are more frequent in spiral arms, with a higher star formation rate. Their effects are more important at large galactocentric radii with a lower vertical component of gravity. With these high values of  $v'$  we would obtain  $\beta$  (outer) larger than =  $10^{28}$  cm<sup>2</sup>s<sup>-1</sup> [93] following recent developments of the fountain model.

The diffusion time  $\tau = l^2/\beta$  becomes  $\tau$  (inner)  $\approx 3 \times 10^9$  years and  $\tau$  (outer)  $\approx 3 \times 10^7$  years. The first one is larger than the rotation time, then the extragalactic  $B_r$  field could not effectively penetrate in the inner disk, but  $B_r$  should have the same extragalactic value in the warped region. Therefore, radial component of the vertical field gradient should be very effective in producing warps.

The magnetic model was designed to explain S-warps (m=1) which is the most frequent type (68% in the large sample of 325 edge-on galaxies [79]). This model barely explains m=0 warps (U-shaped profile); they should be interpreted as arising from gradients in the extragalactic field, with characteristic length of the order of the galaxy size. The generation of asymmetric warps by other mechanisms have been considered [94, 84].

There seems to be a common orientation of warps in relative low-scale clusters. For example it was noticed [96] that in the Local Group the warps of the Milky Way, M31 and M33 show a very noticeable alignment. Also a common orientation of warps in the neighborhood of the Milky Way was found [87, 97]. This fact suggests an extragalactic origin.

It is interesting to notice too that warps are observed in high redshift galaxies [98]. The warps are even larger in these first steps in galaxy evolution. If warps are driven by extragalactic magnetic fields this fact would be interpreted as a result of past higher magnetic strengths. It should be taken into account that due to expansion

$Ba^2 = \text{constant}$ , being  $a$  the cosmological scale factor,  $a^{-1} = z + 1$ , therefore at  $z \sim 1$ , in the Hubble Deep Field, the magnetic strengths were 4 times higher than today.

Intergalactic magnetic fields may produce warps, being the angle of  $45^\circ$  between the rotation axis and the direction of the field, the angle of higher efficiency for bending the disk. When both directions are closer, a larger flaring of the disc would be expected. When the field direction lies close to the galactic plane it would produce lopsidedness, i.e. elliptical shapes in the outer isophotes with radial increase of eccentricity [99]. This could provide a basis of interpretation of observed lopsidedness ([100, 101, 102, 103]). In particular, it was found [101] found that about 50% of spiral galaxies exhibit lopsidedness and that about 20% of the galaxies have a lopsidedness magnitude  $\langle A_1/A_0 \rangle \geq 0.19$  [102]. Also the lopsidedness of the young stellar population is greater than the old one [104]. This indicates that the gas is the first component suffering the distortion. An extragalactic magnetic field is therefore a tentative explanation of lopsidedness that should be explored.

## 6 Galactic magnetic fields.

A very promising field is the interpretation of the CMB taking into account primordial pre-recombination magnetic fields and even the possibility of finding observational traces of their existence with the great sensitivity of the space mission Planck. There is a wide Planck team interested on this topic [107]. A knowledge of primordial magnetic fields provides an understanding of their influence in galaxy birth and evolution.

Observations of CMB can not only provide information about the magnetized pre-galactic medium. They also constitute the most complete direct information of the large scale magnetic field of the Milky Way and some other nearby galaxies. From the all-sky maps obtained at different wavelengths we may obtain the maps of the polarized synchrotron continuum emission and hence very important restrictions on the 3D-distribution of the galactic field.

Preliminary calculations can be obtained using WMAP data. The procedure is as follows: We can select different models proposed in the literature with different free parameters. We then obtain for each model and for each set of parameters of each model the best fit to the real data. We have considered five models and about 70000 sets of parameters per model. In Figure 8 we plot the polarization angle obtained at 22 GHz by WMAP and in Figure 9 the best fit model with the best set of parameters. The model coincides with that proposed in [108] even if different parameters better reproduce the observational map. Nevertheless, this type of approach does not permit the firm withdrawal of other models and we will must wait for Planck, with its much higher sensitivity for polarization to determine the 3D distribution of magnetic fields in our galaxy.

Other techniques [109, 110] to determine the large scale configuration of the

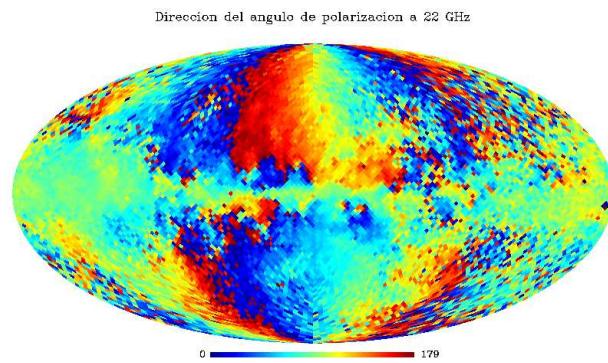


Figure 8: All sky polarization angle from WMAP at 22 GHz.

galactic magnetic field are based on Faraday Rotation of pulsars and extragalactic radio sources. Pulsars inform about the field near the plane and have the advantage of a null intrinsic Faraday Rotation. These techniques do not provide a common picture yet. Probably, the magnetic field of the Milky Way has an “axisymmetric spiral” distribution, distorted by spiral arms and random fluctuations.

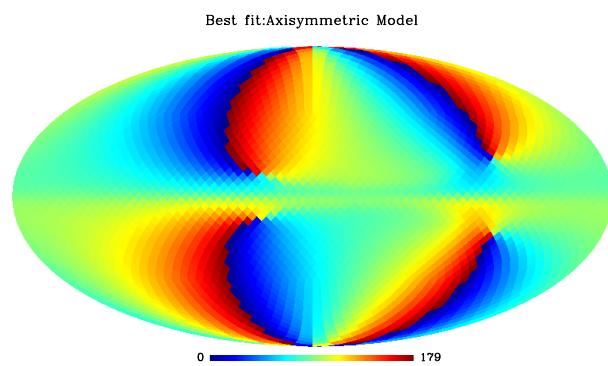


Figure 9: Best fit model of the all sky polarization angle.

## References

- [1] Valleeé. J. 2004, *New Astronomy* 48, 763
- [2] Widrow, L.M. 2002, *Rev. Modern Phys.* 74, 775
- [3] Camenzind, M. 2005, in *Cosmic Magnetic Fields, Lectures Notes in Physics*, Springer, Ed. by R. Wielebinski and R. Beck
- [4] Battaner, E., Isern, J., Florido, E. 1989, *Astrophys. Space Sci.* 162, 35
- [5] Birk, G.T., Lesch, H., Neukirch, T. 1998, *MNRAS* 296, 163
- [6] Dolag, K. 2006, *Astron. Nach.* 327, 575
- [7] Niklas, S. 1995, PhD. Thesis, Univ. Bonn.
- [8] Beck, R. 2005, in “Cosmic Magnetic Fields”, *Lectures Notes in Physics*, Springer, Ed. by R. Wielebinski and R. Beck
- [9] Klein, V., Wielebinski, R., Morsi, H.W. 1998, *A&A* 190, 41
- [10] Beck, R. 2007, *A&A* 471, 93
- [11] Han, J.L., Beck, R., Ehle, M. et al. 1999, *A&A* 348, 405
- [12] Strong, A.W., Moskalenko, I.V., Reimer, O. 2000, *ApJ* 537, 763
- [13] Beck, R. 2001, *Sp. Sci. Rev.* 99, 243
- [14] Beck, R. 2004, *Ap&SSci* 289, 293
- [15] Kim, K.T., Kronberg, P.P., Dewdney, P.E., Landecker, T.L. 1990, *ApJ* 355, 29
- [16] Kronberg, P.P. 2005, in “Cosmic Magnetic Fields”, *Lectures Notes in Physics*, Springer, Ed. by R. Wielebinski and R. Beck
- [17] Feretti, L., Dallacas, D., Giovannini, G. et al. 1995, *A&A* 302, 680
- [18] Kim, K.T., Tribble, P.C., Kronberg, P.P. 1991, *ApJ* 379, 80
- [19] Clarke, T.E., Kronberg, P.P., Böhringer, H. 2001, *ApJ* 547, L111
- [20] Vogt, C., Ensslin, T.A. 2003, *A&A* 412, 373
- [21] Kronberg, P.P. 1994, *Rep. Prog. Phys.* 57, 325
- [22] Battaner, E., Florido, E., Guijarro, A. 2001, in “Galaxy Disks and disks galaxies” *ASP Conference Series* 230,169, Ed. J.G. Funes, E.M. Corsini
- [23] Sofue, Y., Rubin, V. 2001, *ARAA* 39,137
- [24] Nelson, A.H. 1988, *MNRAS* 233,115
- [25] Peratt, A.L. 1988, *Laser and Particle Beams* 6, 471
- [26] Battaner, E., Garrido, J.L., Membrado, M. et al. 1992, *Nature* 360, 652
- [27] Binney, J. 1992, *Nature* 360, 624
- [28] Cuddeford, P., Binney, J. 1993, *Nature* 365, 20
- [29] Battaner, E., Florido, E. 1995, *MNRAS* 277, 1129
- [30] Sánchez-Salcedo, F.J., Reyes-Ruiz, M. 2004, *ApJ* 607, 247
- [31] Battaner, E., Florido, E. 2007, *Astron. Nach.* 328, 92
- [32] Battaner, E., Florido, E. 2000, *Fund. Cosmic Phys.* 21,1
- [33] Navarro, J., Steinmetz, M. 2000, *ApJ* 538, 477
- [34] Hayashi, E., Navarro, J.F. 2002, *BAAS* 34, 1270
- [35] Abadi, M.G., Navarro, F., Steinmetz, M. et al. 2003, *ApJ* 591, 499
- [36] Kutchera, M., Jalocha, J. 2004, *AcPPB* 35, 2493
- [37] Romanovsky, A.J. et al. 2003, *Science* 301, 1696

- [38] Swaters, R. 1999, PhD Thesis, Univ. Gröningen
- [39] Battaner, E., Florido, E., Guijarro, A. 2001, in “Galaxy disks and disk galaxies”, Ed. by J.G. Funes and E.M. Corsini, ASP Conference Series 230, 169
- [40] Chyzy, K.T. et al. 2003, A&A 405, 513
- [41] Nishikori, H., Machida, M., Matsumoto, R. 2006, ApJ 641, 862
- [42] Battaner, E., Lesch, H., Florido, E. 1999, An. Fis. 94, 98
- [43] Sofue, Y., Fujimoto, M., Wielebinski, R. 1986, ARAA 24,459
- [44] Shukurov, A. 2005, in “Cosmic Magnetic Fields”, Lectures Notes in Physics, Springer, Ed. by R. Wielebinski and R. Beck
- [45] Kulsrud, R.M. 2005, in “Cosmic Magnetic Fields”, Lectures Notes in Physics, Springer, Ed. by R. Wielebinski and R. Beck
- [46] van der Kruit, P.C. 1979, AAS 38,15
- [47] van der Kruit, P.C., Searle, L. 1981a, A&A 95, 105
- [48] van der Kruit, P.C., Searle, L. 1981b, A&A 95, 116
- [49] van der Kruit, P.C., Searle, L. 1982, A&A 110, 61
- [50] Barteldrees, A., Dettmar, R.J. 1994, AAS 103, 475
- [51] de Grijs, R., Kregel, M., Wesson, K.H. 2001, MNRAS 324, 1074
- [52] Wainscoat, R., Cohen, M., Völk, K. et al. 1992, ApJSS 83, 111
- [53] Binney, J., Tremaine, S. 1987, in “Galactic Dynamics”, Princeton Series on Astrophysics.
- [54] van der Kruit, P.C. 2008, in “Formation and evolution of Galaxy Disks”, ASP Conference Series Ed. J.G. Funes and E.M. Corsini
- [55] Erwin, P., Beckman, J.E., Pohlen, M. 2005, ApJ 626, L81
- [56] Erwin, P., Pohlen, M., Beckman, J.E. 2008, AJ 135, 20
- [57] Pohlen, M., Trujillo, I. 2006, A&A 454, 759
- [58] Trujillo, I., Pohlen, M. 2005, ApJ 630, L17
- [59] Pohlen, M., Zaroubi, S., Peletier, R. et al. 2007, MNRAS 378, 594
- [60] Florido, E., Battaner, E., Guijarro, A. et al. 2001, A&A 378, 82
- [61] Florido, E., Battaner, E., Guijarro, A. et al. 2006a, A&A 455, 467
- [62] Florido, E., Battaner, E., Guijarro, A. et al. 2006b, A&A 455, 475
- [63] Florido, E., Battaner, E., Zurita, A. et al. 2007, A&A 472, L39
- [64] Battaner, E., Florido, E., Jiménez-Vicente, J. 2002, A&A 388, 213
- [65] Kennicutt, R.C. 1989, ApJ 344, 685
- [66] Schaye, J. 2004, ApJ 609, 667
- [67] Fall, S.M., Efstathiou, G. 1980, MNRAS 193, 189
- [68] van der Kruit, P.C. 2007, A&A 466, 883
- [69] Battaglia, G., Fraternali, F., Oosterloo, T. 2006, A&A 447, 49
- [70] Sánchez-Saavedra, M.L., Battaner, E., Florido, E. 1990, MNRAS 246, 458
- [71] Reshetnikov, V., Combes, F. 1998, A&A 337, 9
- [72] Sánchez-Saavedra, M.L., Battaner, E., Guijarro, A. et al. 2003, A&A 399, 457
- [73] Reshetnikov, V., Battaner, E., Combes, F. et al. 2002, A&A 382, 513
- [74] Pérez, I. 2004, A&A 427, 47
- [75] Sancisi, R. 1976, A&A 53, 159

- [76] Briggs, F.H. 1990, ApJ 352, 15
- [77] García-Ruiz, J., Sancisi, R., Kuijken, K. 2002, A&A 394, 769
- [78] Florido, E., Prieto, M., Battaner, E. et al. 1991, A&A 242, 301
- [79] Ann, H.B., Park, J.C. 2006, NewA 11, 293
- [80] Binney, J.J. 1992, ARAA 30, 51
- [81] Battaner, E., Florido, E., Jiménez-Vicente, J. et al. 1997 in “The impact of large scale near-IR surveys”, 49, Ed. F.Garzón. Kluwer
- [82] Weinberg, M.D., Blitz, L. 2006, ApJ 641L, 33
- [83] Sánchez-Salcedo, F.J. 2006, MNRAS 365, 555
- [84] López-Corredoira, M., Betancort-Rijo, J.A., Beckman, J.E. 2002, A&A 386, 169
- [85] López-Corredoira, M., Betancort-Rijo, J.A., Beckman, J.E. 2008, astro-ph/07061550, in “Pathways through an eclectic Universe”, Ed. by J. Knappen, T.J. Mahoney, A. Vazdekis, ASP Conference Series.
- [86] Kahn, F.D., Woltjer, L. 1959, AJ 130, 705
- [87] Battaner, E., Florido, E., Sánchez-Saavedra, M.L. 1990, A&A 236, 1
- [88] Battaner, E. 1996, in “Astrophysical Fluid Dynamics”, Cambridge Univ. Press
- [89] Ruzmaikin, A.A., Sokoloff, D., Shukurov, A.M. 1989, MNRAS 241, 1
- [90] Sokoloff, D., Shukurov, A.M. 1990, Nature 347, 51
- [91] Ruzmaikin, A.A., Shukurov, A.M., Sokoloff, D.D. 1988, in “Magnetic fields in galaxies”, Kluwer Ac.
- [92] Norman, C., Ikeuchi, S. 1989, ApJ 345, 372
- [93] Avillez, M.A., Breitschwerdt, D. 2004, A&A 425, 899
- [94] Saha, K., Jog, C.J. 2006, A&A 446, 897
- [95] Sparke, L., Casertano, S. 1988, MNRAS 234, 873
- [96] Zurita, A., Battaner, E. 1997, A&A 322, 86
- [97] Battaner, E., Garrido, J.L., Sánchez-Saavedra, M.L. et al. 1991, A&A 251, 402
- [98] Reshetnikov, V., Battaner, E., Combes, F. et al. 2002, A&A 382, 513
- [99] Battaner, E., Florido, E., Sánchez-Saavedra, M.L. 1992, A&A 253, 89
- [100] Baldwin, J.E., Lynden-Bell, D., Sancisi, R. 1980, MNRAS 193, 313
- [101] Richter, O.G., Sancisi, R. 1994, A&A 290, L9
- [102] Rudnick, G., Rix, H.W. 1998, Astron.J 116, 1163
- [103] Angiras, R.A., Jog, C.J., Omar, A. et al. 2006, MNRAS 369, 1849
- [104] Heller, A.B., Brosch, N., Almoznino, E. et al. 2000, MNRAS 316, 569
- [105] Chyzy, K.T., Beck, R., Kohle, S. et al. 2000, A&A 355, 128
- [106] Grasso, D., Rubenstein, H.R. 2001, Phys. Rep. 348, 163
- [107] Battaner, E. et al. 2007, poster in PLANCK Toulouse Meeting
- [108] Page, L. et al. 2007, ApJSS 170, 335
- [109] Wielebinski, R. 2005, in “Cosmic Magnetic Fields”, Lectures Notes in Physics, Springer, Ed. by R. Wielebinski and R. Beck
- [110] Wolleben, M., Landecker, T.L., Reich, W. et al. 2006, A&A 448, 411

# GRAN TELESCOPIO CANARIAS: A KEY ASSET FOR SPANISH ASTRONOMY

FRANCISCO SÁNCHEZ MARTÍNEZ

*Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, SPAIN*

**Keywords:** Telescopes – Instrumentation – Observatories

First of all I would like to thank the organisers of this Astrophysics Symposium (Ana Ulla and Minia Manteiga) for inviting me to take part, and congratulate them on a very successful event.

I would like to begin by stating the obvious: the history of Astronomy is the history of astronomical observation. We should remember that the reasons for which it has been practiced have changed throughout human evolution. In early times it was used to predict the future (astrology) and for mythic and ritual purposes (primitive religion). Practical applications, such as agriculture, navigation and new technologies, have always been sought for it and in very recent times it has been used in mathematics (celestial mechanics) and physics (Astrophysics). With these latest steps, Astronomy has completed its evolution: Understanding the Universe (including LIFE) both as a whole and in all its detail is now the sole aim of our research into space.

Observation will continue to be the key to astronomical knowledge. All recent studies into the future of Astronomy have confirmed this, and they have also pointed to the need for giant telescopes to be installed at the world's best sites. Spain is lucky in this respect as the Canaries are one such location.

All of this is relevant to understanding why the Gran Telescopio CANARIAS (GTC) was built and the impact it will have in the future. This is my subject today.

Three things inspired Spain to embark on this daring project to build what is the largest and most advanced optical infrared telescope of its time: the flourishing Astrophysics community in Spain; the level of technical ability achieved by the country's industry; and the excellent observatories in the Canaries.

Anyone involved in characterising observation sites around the world will know that the astronomical quality of the observatories in the Canaries has been more thoroughly researched than any other, and that the reams of data generated by that research testify to the permanent nature of their extraordinary quality. In excess of seventy institutions from nineteen different countries have installed instruments at these observatories, which belong to the Instituto de Astrofísica de Canarias (IAC), and together they make up the European Northern Observatory.

From the beginning of the project, the Gran Telescopio CANARIAS has had the following guiding aims:

1. To provide the Spanish Astronomy community with its own cutting-edge instrument:
  - to involve the community from the start (feasibility study, design, construction);
  - focal instruments specifically designed to meet the needs of the Astrophysics community;
  - to prepare observation programmes very early (this is why the three international symposiums on *Science with the GTC* have been held);
  - to encourage our research centres to participate in developing the instruments.
2. To ensure that the observatories in the Canaries continue to be amongst the world's leading observatories:
  - to lead (Spain) the design and construction of the GTC;
  - to site the telescope at the Observatorio del Roque de los Muchachos;
  - to continue ongoing characterisation of the skies over the Canaries;
  - to strengthen enforcement of what is known as the “Law of the Sky”.
3. To stimulate the development of Spanish industry in the field of advanced technology:
  - to build up the capacity of companies;
  - research centres and industry should tender for work jointly;
  - an open and interactive tender process;
  - to provide information flexibly and on an ongoing basis from the “Project Office”.

Two of these aims have already been met: the observatories in the Canaries are front runners in the contest to host the super giant telescopes of the immediate future, and Spanish companies involved in building the GTC have amassed the knowledge and reputation they will need to secure the most rewarding international contracts. The first objective will not truly be achieved, however, until results produced by our Astrophysics community once the telescope has entered service have propelled some of its members into leading positions in important fields of knowledge.

## 1 Principal features of the GTC

- A large collecting surface: it will have a segmented primary mirror 11.4 meters in diameter (equivalent to a monolithic mirror of 10.4 meters).
- Excellent image quality: the mirrors will be as highly polished and the adaptive optics system will be as accurate as current technology allows.
- From its inception, the telescope has been designed to use adaptive optics.
- High levels of reliability: 2% reductions in time lost to maintenance and breakdowns.
- High levels of operating efficiency.
- Advanced control system.
- Only one secondary mirror (it can work in both the visible and infrared ranges).
- Collapsible tertiary mirror.
- Classical and queued observation.

## 2 “Day One” focal instruments

- OSIRIS (optical), for spectroscopy and imaging.
- CANARICAM (infrared), for spectroscopy and imaging.
- ELMER (optical), substitute instrument for OSIRIS.

## 3 Second generation focal instruments

- EMIR: near-infrared (0.9 to 5 microns) imager and spectrograph.
- FRIDA: spectroscopy and imaging with adaptive optics.

## 4 Third generation focal instruments

- SIDE: multiple spectrometer with variable resolution.
- Many other instruments yet to be developed.

Obtaining this powerful tool for what is known as “big science” has been very laborious. For an understanding of the work involved, the following chronology is instructive. Even this, though, cannot truly portray the problems we had with the most serious obstacle of all: the culture of mistrust and jealousy that had to be overcome both in Spain and abroad.

1987. The William Herschel Telescope (4.2 m) enters service at the ORM:

- British (RGO) and Spanish (IAC) scientists start work to design an 8 meter class telescope.

1989. Preparations for an 8 metre monolithic telescope are completed:

- a presentation is made to the Governing Council of the IAC and the CCI (International Scientific Committee) of the Canarian observatories, and is received very positively
- formal negotiations begin between Spain and the UK to jointly build the telescope.

1990. In Britain, the SERC decides to abandon the project, opting instead for the American Gemini project:

- the IAC decides to continue with the project alone.

1993. Funds are secured from FEDER to undertake studies into the possibility of building a large telescope in the Canaries:

- consultation with the Spanish astronomical community confirms that it is in favour of the project.

1994. The Government of the Canaries (through the express wish of President M. Hernaldo) sets up GRANTECAN S.A. to build the GTC, a turning point in the project:

- the feasibility study and pilot project for a monolithic 8 m telescope begins.

1995. Meeting of experts on large telescopes is held at the IAC to scrutinise the proposal:

- at the last minute the experts are consulted on the potential for building a segmented 10 metre class telescope. They are of the opinion that this is the future
- the existing feasibility study is considered fit for the new concept.

1996. The Governing Council of the IAC (under the Chairmanship of Minister J. Saavedra) approves the viability study and gives the project the “green light”:

- the Spanish Government makes its involvement conditional on the necessary funds being available in advance, on three other countries participating, and on the project being overseen by a foreign expert in large telescopes (following the recommendations of the Evaluation Committee for Large Science Installations)
- the “demonstration contracts” designed to build capacity for the project in Spanish Industry are in place (with funding from the national Ministry of Industry and the Department of Industry of the Canarian Government)
- Spain’s membership of ESO ‘to extend into the Northern Hemisphere from the Canaries,’ is about to be signed. The discussions break off as a result of the change of Government in Spain, just as the possibility of installing a 5th VLT at La Palma, and the participation of the ESO in building the GTC as a means of payment for Spain’s membership of that organisation, are being investigated.

1997. The conceptual design of the GTC is completed, and work is underway to secure the funds needed to build it.

1998. The Office for Science and Technology (on the instruction of its Director F. Aldana) succeeds in unblocking the project:

- the Spanish Government, from this time onwards, becomes a full participant in the project and leads the GRANTECAN S.A. public company.

1999. The GTC is mentioned explicitly in the State Budgets, and in the FEDER2000.

2000. His Royal Highness the Prince of Asturias lays the “foundation stone” at the Observatorio del Roque de los Muchachos.

2001. Mexico (Universidad Nacional Autónoma de México, and the Instituto Nacional de Astrofísica Óptica y Electrónica) and the United States (University of Florida) become partners.

2001. The “National Programme of Astronomy and Astrophysics” begins (as part of the national Research and Development Plan), as a result of the GTC project.

2002. The “civil engineering works” are completed and the handover of the dome takes place.

2003. Assembly of the telescope structure begins at the Observatory.

2004. The first segments of the primary mirror are received.

2005. The framework for the telescope is completed and delivery is taken of the tertiary mirror and Acquisition and Guidance camera.

2007. “First Light” at the GTC.

The GTC is currently on the way to being completed (its “First Light” on the 13th July 2007 was a notable success), although many details remain to be finalised before it is fully operational. *Within a year it will be open for use by the astronomical community*, with “day one” instruments (OSIRIS + CANARICAM), and substitute instrument (ELMER). Observing time will be available in the “008 B” semester (October 2008 - March 2009). “ESO time” (time given over to the European Southern Observatory as part payment for Spain’s membership of that organisation) will be available from January 2009.

I will never tire of congratulating all of the people responsible for bringing the project to fruition, headed by Pedro Álvarez (Project Director) and José Miguel Rodríguez Espinosa (Scientific Manager). Over time their tremendous contribution will come to be valued even more.

The time has come to begin producing front-line science with this great telescope. With its large collecting surface (75 square metres) the GTC is well-suited to observing very weak objects, like the distant galaxies that show us the origins of the stars which populate the Universe. It also brings advances for observing faint objects, and for investigating the birth and death of stars, quasi-stellar objects and planets. The large diameter of its primary mirror means that it will see fine detail, in near stars, other galaxies, in the structures of protostellar clouds, star clusters and more. Its capabilities in the infrared range will also allow it to penetrate dense clouds of gas and dust.

We hope, too, that new and unexpected discoveries will be made before long. Could they even be so significant that they will change our understanding of the Universe?

In addition to all this, once the GTC is in operation it will be a great ambassador for the development of technology and industry in Spain. In the future, it will help us secure the siting of the “European ELT” and the EST (European large Solar Telescope) in Spain, together with a role for our research centres and companies in their construction.

We also aim to use the GTC to train researchers and technologists and to promote science to the general public. The prospects could not be more exciting, especially for the new generations of our astrophysicists.

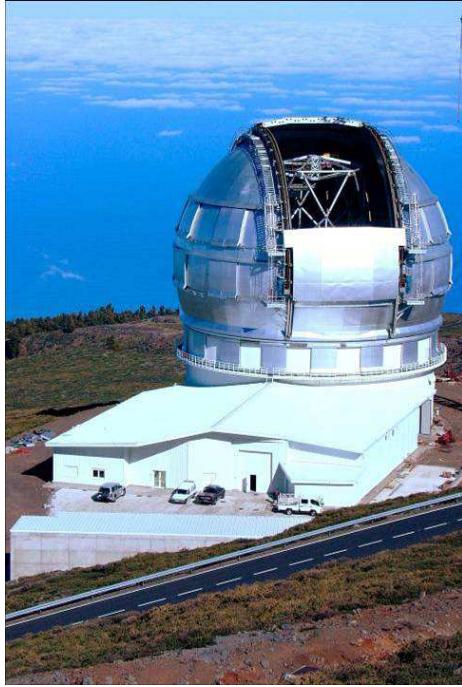


Figure 1: Exterior view of the Gran Telescopio CANARIAS (GTC), installed at the IAC's Observatorio del Roque de los Muchachos.



Figure 2: View of the GTC from inside the dome.

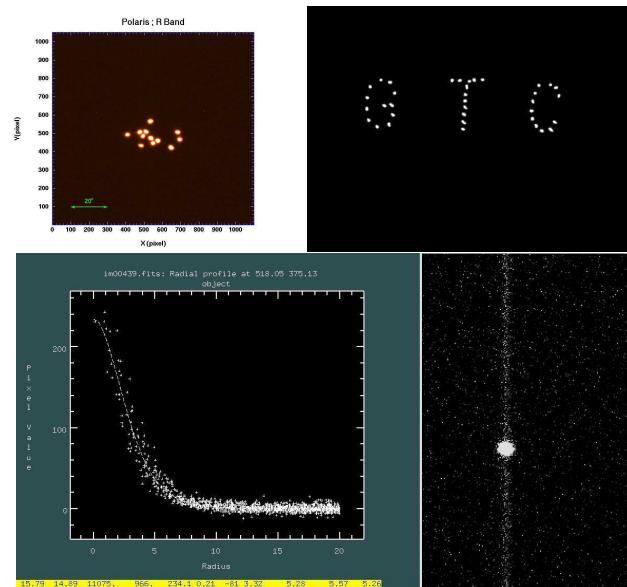


Figure 3: First light. Upper left panel: when the images of the star observed (one from each of the primary mirror segments) were still separated. Upper right panel: The images of the observed star arranged to form the letters GTC. This demonstrates the mastery over mirror adjustment that already exists. Lower panel: The star observed (the images from each primary segment are “stacked” here), showing the extraordinary resolution achieved, without optical adjustment.



Figure 4: First Light: The principal assistants drinking a Ltoast in the telescope control room immediately after achieving a very successful “first light”. Happiness is written on every face.

# HIGH-RESOLUTION GROUND-BASED EUROPEAN SOLAR PHYSICS

MANUEL COLLADOS

*Instituto de Astrofísica de Canarias, E-38205, La Laguna, SPAIN*  
and THE EST TEAM

**Abstract:** This communication reviews some of the most challenging topics in high-resolution ground-based Solar Physics. The most powerful European facilities are described, together with their capabilities and skills gained in Europe using them. The reasons for a large-aperture solar telescope are outlined, based on present scientific needs, which have led to the joint project EST (European Solar Telescope), in which the most prestigious European Solar Physics research institutions participate. Some technical challenges of a such a large telescope are mentioned.

**Keywords:** Sun – Solar Telescopes – Instrumentation – High Resolution – Magnetic Fields

## 1 Introduction: The Sun as a three-fold star

Understanding the processes that take place in our Sun is crucial for many reasons. Firstly, there is a fundamental *Solar-terrestrial connection*. The Sun is of paramount importance because it sustains life on Earth. Changes in its conditions may have dramatic consequences for us. Large amounts of energy, derived from the magnetic energy stored in the fields, are deposited into the plasma on very short timescales of seconds to minutes. These flares can accelerate plasma to velocities up to a significant fraction of the speed of light, and if the bulk of the accelerated plasma (in the form of a coronal mass ejection) hits the magnetopause of the earth, it creates fascinating events (Aurorae) and also potentially hazardous phenomena for our living environment (damage to satellites, overload of power lines, increased radiation dose for aircraft/ISS crew, etc.). It is thus fundamental to study these processes in order to be able to predict them. The question of how both long and short-term changes in the solar irradiance can affect the terrestrial climate system is also clearly one of great importance for the accurate prediction of future trends in global warming. To do this requires a much clearer understanding of how changes in the magnetic field affect the solar irradiance than we currently have through the combined efforts of observation and modelling. However, the characteristic spatial and temporal scales of

the physical processes in the solar atmosphere are often determined by the magnetic field and are typically so small that their study is impossible due to lack of photons. Large collecting areas are absolutely necessary for a final understanding of the physics of magnetised solar plasma.

Secondly, there is the *planetary and astrophysical connection*. The Sun has a direct influence on all the planets in the heliosphere and there are many parallels between the processes that occur in the solar atmosphere and in planetary magnetospheres, including our own. The Sun is a star and has been used many times to guide stellar evolution models. Many stars are known to have magnetic cycles similar to that of the Sun, and starspots are beginning to be detected thanks to special techniques like Doppler imaging. The feedback between solar and stellar observations is crucial in allowing us to determine how our Sun will evolve in the future. Magnetic fields are also of critical importance in understanding most astrophysical systems, e.g. in the formation of jets observed in many astrophysical objects. They may also play an important role in galaxy evolution. These are only a few examples in which solar studies can help in understanding our universe.

Finally, there is a question of *basic science*. The Sun is the only place where we can study in detail the interaction of plasma and magnetic field, and, as such, it can be considered as a fundamental physics laboratory. Because of its tremendous size, magnetic diffusivity times are very large making it possible to follow the evolution of this interaction in a way that is impossible to reproduce in a laboratory. Consequently, the Sun is the only place where phenomena like conversion of mechanical energy into magnetic energy, conversion of magnetic energy into thermal energy, or particle kinetic energy can be studied by direct observation and in detail. Our Sun is unique in that it is currently the only star with a surface that can be spatially resolved at a level approaching that at which physical processes occur.

## 2 Present situation of ground-based high-resolution European solar physics

During the last 20 years, a number of European countries have constructed powerful ground-based telescopes that have increased our knowledge of the Sun. The most advanced European facilities are presently located in the Canary Islands observatories. Table 1.1 lists the nowadays existing, or foreseen for a very near future, European high-resolution solar telescopes, together with some of their most powerful capabilities. They lie in the 0.5-1.5 metre range.

On the basis of their existing observing facilities, European solar astronomers have gained leadership positions in several areas of instrumentation and data analysis. These areas include high resolution bi-dimensional spectroscopy (Italy and Germany), polarimetry at visible and infrared wavelengths (France, Spain, Italy, and Switzer-

Facility (Location)	Aperture [m]	Wavelength range ( $\mu\text{m}$ )	Capabilities
Dutch Open Telescope (La Palma, Spain)	0.45	0.35 – 0.8	High resolution imaging
Vacuum Tower Telescope (Tenerife, Spain)	0.7	0.35 – 2.3	High resolution imaging Adaptive optics Spectropolarimetry
THÉMIS (Tenerife, Spain)	0.9	0.40 – 1.2	Spectropolarimetry
Swedish Solar Telescope (La Palma, Spain)	1.0	0.35 – 1.1	High resolution imaging Adaptive optics Spectropolarimetry
GREGOR (Tenerife, Spain)	1.5	0.35 – 12	High resolution imaging Adaptive optics Spectropolarimetry

Table 1: European high resolution research telescopes and their capabilities.

land), adaptive optics development (Sweden and Germany), and post-processing techniques for diffraction-limited imaging (Sweden, Germany and The Netherlands).

High-precision visible and near-infrared spectropolarimetry is now routinely performed at the German Vacuum Tower Telescope (VTT) and at the French-Italian THÉMIS telescope. Accurate spectropolarimetry ( $10^{-3}$  sensitivity or better) is very photon demanding and integration times of at least several seconds are required to detect the necessary number of photons to reduce noise to acceptable levels. Only very recently, with the development of powerful adaptive optics systems has it been possible to increase the spatial resolution of spectropolarimetric data. But, even today, observations with a spatial resolution better than 0.3-0.5 arcsec are scarce. Improving this situation will only be possible if a strong effort is made to increase the telescope aperture, thus making possible the detection of a larger number of photons from smaller solar areas, complemented with a multi-adaptive conjugate adaptive system to increase the size of the corrected field of view (see Fig. 1).

Near-infrared observations offer several advantages over standard visible techniques. One of them is their larger magnetic sensitivity. Magnetic fields are better detected and studied in this wavelength range. A second argument in favour of infrared observations is the reduced sensitivity to atmospheric disturbances. However, due to the dependence of the diffraction limit on wavelength, a larger telescope is required to observe the Sun in the near-infrared with a spatial resolution of, say, 0.1 arcsec.

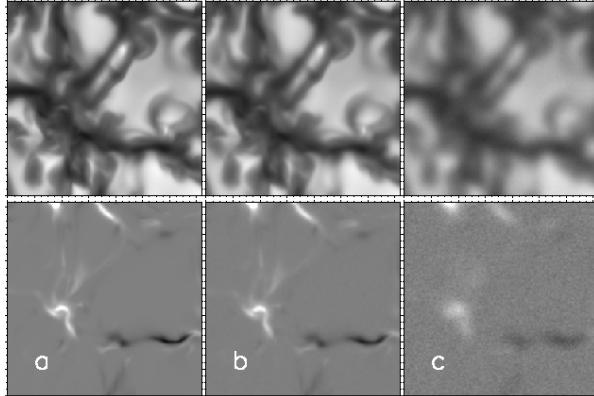


Figure 1: *a*) A frame taken from a 3-D simulation (20 km grid) of magneto-convection near the solar surface (courtesy of E. Khomenko). Each box is 3000 km (4 arcsec) wide and shows a snapshot of the intensity (top) and vertical magnetic field strength (bottom) as it would appear looking down on the simulation. *b*) The same areas seen through a diffraction-limited 4m telescope at a wavelength of 630 nm (a noise of  $10^{-3}$  has been added to simulate an observation in a realistic way) and, *c*) Same images as seen with a 0.75m telescope at the same wavelength (noise has been added as if the images had been taken with the same integration time as in *b*). The model predicts rapidly evolving, highly mixed, bipolar fields that are twisted by strong turbulent downflows. A 4-meter class telescope is required to measure both the spatial and temporal characteristic of the magnetic field associated with this small-scale dynamo process which current telescopes cannot detect and resolve.

Numerical magnetohydrodynamical simulations presently achieve a degree of complexity that allows direct comparison with observations. In fact, numerical spatial resolutions, of the order of 10 km, are better than the observations and the simulations need to be degraded for an adequate comparison. The interaction of convective flows with magnetic fields has been reproduced, showing how plasma motions sweep the magnetic field to converging flow regions (downflows) where the field is concentrated and intensified. Small-scale magnetic field structures channel energy and seem to be the fundamental key for understanding how the chromosphere and corona are heated. Simulations, together with high-spatial resolution observations, will ultimately reveal e.g. the origin of the energy deficit in the umbra of sunspots or the energy excess in sunspot penumbrae. The spatial resolution of observational data is today signifi-

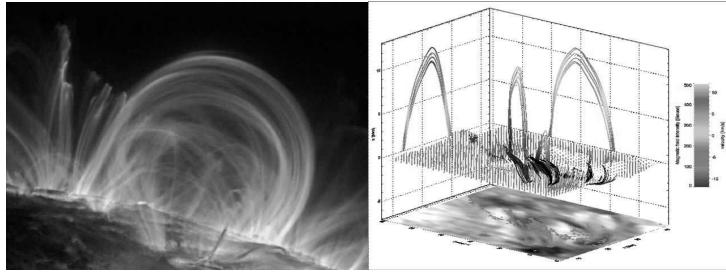


Figure 2: *Left:* Loops observed in the solar corona by the satellite TRACE. *Right:* Magnetic field configuration of an emerging region, showing similar loops as those observed by TRACE. These results have been obtained using near-infrared spectropolarimetric data taken at the German VTT.

cantly poorer than that possible with numerical simulations, preventing progress in our understanding of, e.g., the internal structure of thin flux tubes or the propagation of different MHD waves in magnetic structures. Simulations will undoubtedly continue to improve, thanks to increasingly powerful computers and better modelling of radiative transfer. Better quality data than currently available is then essential to observe the phenomena that are only accessible today via numerical calculations. Unexpected findings, when increasing the spatial and temporal resolution of observational data, are to be expected, which will in turn also serve as input for improving simulations. This has been the case, for instance, with the discovery of penumbral dark cores, which were only discovered when observations with a spatial resolution of 0.1 arcsec were achieved. Their magnetic properties remain unknown, however, because of the inherent difficulty of reaching such a resolution with currently available spectropolarimetric data. There will certainly be a mutual benefit between numerical simulations and high quality observations obtained with a large aperture telescope.

The theory of radiative transfer in a magnetised medium was developed several decades ago. The formation of spectral lines under LTE conditions giving rise to polarised spectral lines via the Zeeman effect is very well understood. Inversion codes exist that retrieve the solar atmospheric conditions under which the spectral lines are formed. Temperature, velocity and magnetic field stratifications are routinely calculated in the photosphere. However, only now are we starting to understand and reproduce the formation of spectral lines in the chromosphere, where non-LTE conditions prevail (see Fig. 2). The Hanle effect has also been demonstrated to leave its imprint on spectral lines, and extremely weak magnetic fields can be detected with it. The foundations are now in place for the development of the tools required

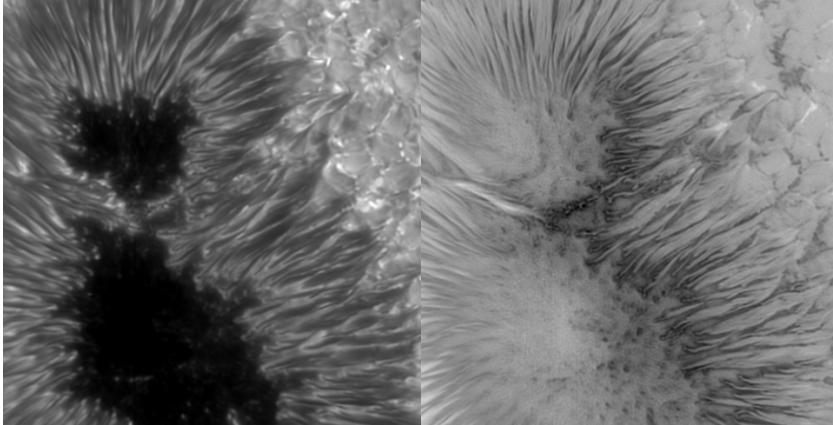


Figure 3: Sunspot recorded by the SST in the blue and red wings of the magnetically sensitive Fe I line at 630.2 nm on Sept 12, 2006. The left-hand image represents the average intensity, the right-hand image the difference in circular polarization (normalised to the intensity) of the two wings. The spatial resolution of these images, consisting of over 500 individual exposures, is the result of adaptive optics and image restoration techniques (developed by Löfdahl and van Noort) and is close to  $0''.2$ . Dark penumbra cores, discovered with the SST, are seen clearly in both intensity and polarization signal (Michiel van Noort, Institute for Solar Physics)

to diagnose the magnetic field in the solar chromosphere. These will be a fundamental key for the correct interpretation of observational data corresponding to this highly dynamic layer.

The 1-metre Swedish Solar Telescope (SST) presently produces the sharpest images of the Sun ever obtained (see Figs. 3 and 4). Not even space missions, free from atmospheric turbulence, can reach that degree of detail, because of the smaller size of the telescopes that can be launched. Such diffraction-limited observations require excellent optics and effective adaptive optics working on an extended object to counter the disturbances in the Earth's atmosphere. Post-facto reconstructions also help reach this high image quality. The 45-cm Dutch Open Telescope (DOT) has proven that an open telescope is capable of obtaining excellent results. GREGOR, with its 1.5 metre aperture and open structure, will represent an intermediate step towards a larger telescope and will serve as a test bench for the development of new generation high-order multi-conjugate adaptive optics.

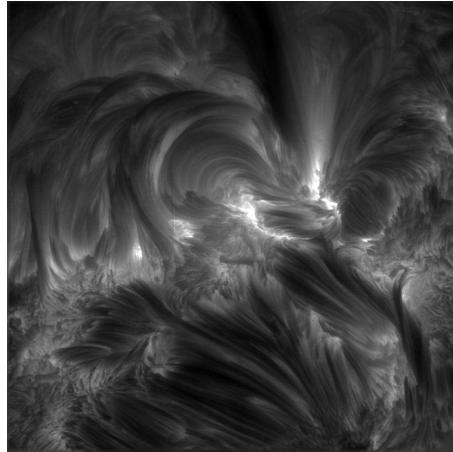


Figure 4:  $H\alpha$  line core image of an active region observed with the SST on 04-Oct-2005. The FOV is  $64 \times 64$  arcsec. The spatial resolution is close to the diffraction limit of the telescope ( $\sim 120$  km) which was achieved with the aid of adaptive optics and post-processing techniques (Multi-object Multi-Frame Blind Deconvolution).

### 3 Future perspective

The aperture of a telescope is the essential characteristic in determining its resolving power, but until recently ground-based solar telescopes have been more limited by the effects of atmospheric distortion that disturb the incoming wavefront. New and more powerful adaptive optics systems installed at some of the principal solar telescopes, including the German Vacuum Tower Telescope and the Swedish Solar Telescope, are now able to overcome a large fraction of the atmospheric distortion. This has not only given us tantalizing glimpses into the very fine scale structures on the solar surface, but has shown that these systems are now so mature that our ability to resolve small detail is now limited by the size of the telescopes themselves, and not by the atmospheric distortions.

Besides spatial resolution, the light collecting power of a large aperture is crucial for solar research, quite contrary to the perception that sunlight is bright. Magnetic fields are detected by measuring the polarisation of magnetically sensitive spectral lines. The fraction of the light in a spectral line that is polarised is tiny (sometimes

below  $10^{-3}$ ). The precision of such measurements is thus fundamentally limited by photon statistics. A larger aperture collects more photons from a given area on the solar surface, permitting the required precision for polarimetric measurements of one part in  $10^4$ . The timescales with which solar features change are related to the speed of sound in the solar atmosphere (about 7 km/s), so smaller features evolve faster than larger ones. The required time resolution amounts to just a few seconds, making a large aperture even more necessary.

In view of all these arguments, solar astronomers world-wide are unanimously in agreement that a much larger step in observational capability is needed to fully understand the fundamental processes of plasma physics that are at work in the outer layers of our star, and, in particular, to begin an in depth study of the magnetic coupling of the solar atmosphere by simultaneously observing the photosphere and the chromosphere, with high spatial resolution (to see details smaller than 100 km), high spectropolarimetric accuracy (for an accurate determination of magnetic fields and velocities) and high temporal resolution (for evolutionary studies at time scales of a few seconds). Numerical simulations and radiative transfer calculations, taking into account the Zeeman and Hanle effects, will help interpret the observational data. A large aperture telescope is thus required to meet these requirements.

Following this trend, many leading European solar astronomers decided to form the consortium EAST (European Association for Solar Telescopes) to keep Europe in the frontier of international solar physics. Among others, one of the aims of this consortium is to develop, construct and operate a next generation large aperture telescope (EST, European Solar Telescope) in the Canaries. This telescope should be optimised for studies of magnetic coupling between the deep photosphere and upper chromosphere. This will require diagnostics of the thermal, dynamic and magnetic properties of the plasma over many scale heights, by using multiple wavelength imaging, spectroscopy and spectropolarimetry. The EST design will therefore strongly emphasise the use of a large number of instruments simultaneously, thereby improving photon efficiency and diagnostic capabilities relative to other existing or proposed ground-based or space-borne solar telescopes. To implement the EST science goals, also high spatial and temporal resolution will be needed.

## 4 Some science cases for EST

Some examples where a large aperture solar telescope is necessary to make significant progress are the following:

- Magnetic flux removal from active regions

The magnetic field generated during one solar cycle must disappear from the surface before the opposite-polarity field of the next cycle can appear, i.e., a

removal of the photospheric field is required for the dynamo to operate. Active regions and sunspots are observed to slowly decay with time: their size and magnetic flux start to decrease as soon as they are fully formed. Part of the flux lost by sunspots migrates to the poles through random walk diffusion, producing the magnetic field reversal observed in the polar caps around the maximum of sunspot activity. The fate of the remaining sunspot flux is unknown.

Magnetic flux can be eliminated from the photosphere by in-situ small-scale reconnection processes or by the rise (submergence) of field lines to higher (deeper) layers. The edge of the moat surrounding sunspots is known to be the site of many flux cancellation events, i.e., the disappearance of flux concentrations of opposite polarity as they come into close contact. Often, moving magnetic features associated with the sunspot cancel when they hit existing plate elements of opposite polarity at the boundary of the moat. It has been suggested that these cancellations may explain the decay of sunspots and hence the removal from the solar surface of a significant fraction of the magnetic flux in active regions. With current or upcoming observations, however, it is difficult to determine whether these small-scale cancellations are the result of magnetic reconnection, the rise of U-loops, or the submergence of  $\Omega$ -loops. All three processes may be a source of chromospheric/coronal heating (due to the release of magnetic energy), and of transient events in the upper atmosphere. To distinguish among the different possibilities, simultaneous observations of the photosphere and chromosphere need to be carried out at the highest angular resolution possible. By analyzing the photospheric and chromospheric flow fields at the cancellation sites, and the timing difference between the events occurring in the different layers, it will be possible to draw definite conclusions about which mechanism is actually responsible for the removal of magnetic flux.

- Flux emergence and cancellation in the quiet sun

The two sources of flux for the network are the dispersal of flux from active regions and ephemeral regions. Ephemeral regions are bipolar magnetic concentrations with sizes typically less than 20000 km that appear over the entire quiet Sun and, subject to granular motions, merge and cancel with other network flux concentrations. They are short-lived, with an average lifetime of 4.4 hours. Through cancellation processes, ephemeral regions replace the total flux of the quiet sun in 8-19 hours. We still do not know if ephemeral regions and internetwork fields are produced by the global dynamo, a local dynamo, or both. To answer this question, it is important to follow their temporal evolution during emergence and to determine their magnetic field topology and connectivity in the different layers of the atmosphere. It has been suggested that, in general, all flux emergences above  $10^{18}$  Maxwell are associated with coronal brightening in Fe XII ( $\sim 1.6$  MK). Emerging loop structures first become visible in the

corona when the separation of the two poles of the ephemeral region reach 4.5-8.5 Mm, and then remain visible for 5-12 h. These results suggest that there is magnetic field connectivity between the photosphere and corona at the location of ephemeral regions, but no further studies have been carried out so far, partly because of the lack of reliable magnetic field measurements in the chromosphere (the interface layer between photosphere and corona). Whether this connectivity is able to provide a heating mechanism to the corona remains unclear.

Flux cancellations are very frequent in the quiet Sun. Many of these events involve one of the poles of an ephemeral region and a network element of opposite polarity. As a result of the cancellation, the network element disappears and is effectively replaced by the other pole of the ephemeral region. The reconfiguration of the photospheric field can result in localised heating in the chromosphere and corona by field braiding. Brightenings in the chromosphere and corona are indeed observed during quiet Sun cancellation events, but it is not clear whether they are caused by field braiding, the retraction of  $\Omega$ -loops, the emergence of U-loops, or photospheric reconnection. EST will clarify this issue by making it possible to investigate the time difference between events in the photosphere and chromosphere, and by providing information about the connectivity of the field in the photosphere and upper layers by direct measurements of the magnetic field. Once this fundamental problem is solved, it will be possible to assess the potential of cancellation events as a source of atmospheric heating in the quiet Sun.

- Wave propagation in different magnetic structures

Waves are one of the most promising candidates for upper atmospheric heating. It has been suggested that acoustic waves are generated by the turbulent convection in the sub-photospheric layers. In a stratified solar atmosphere, an acoustic cut-off frequency exists producing a reflection of the low-frequency waves back to the lower atmosphere and convection zone. Only high-frequency waves above the cut-off frequency can propagate outwards. These high-frequency waves can propagate to the chromosphere, steepen into shocks and dissipate their energy there. However, the acoustic energy flux of these waves has been found to be at least ten times lower than required to balance the radiative losses in the solar chromosphere.

Magnetic fields play an important role in modifying the properties of the acoustic oscillations. In the presence of a magnetic field, a new characteristic wave propagation speed appears: the Alfvén speed. In those layers where both acoustic and Alfvén speeds are similar, the energy of acoustic waves can be effectively transformed into the energy of magneto-acoustic waves. Magneto-acoustic

waves can also be generated by the random motion of the photospheric foot-points of the fields, which are stochastically shuffled around by the surrounding convective flows. Since the fields connect photosphere and chromosphere, they provide channels for energy transport to the upper layers.

However, the efficiency of mode conversion or the excitation of magnetoacoustic waves depends on the size of the magnetic structures. For weak inter-network fields and the magnetic network with field strength up to 1.5 kG, plasma motions determine the field topology, whereas the strong fields of pores and sunspots suppress the convection and direct the flows along their field lines. While the number of sunspots changes with the solar cycle, the small-scale network flux tubes and the internetwork flux are always present on the solar surface. To address the role of magnetic fields in chromospheric heating, one thus has to determine both the relative contribution of the different kind of magnetic structures to the heating, and the actual way of energy transport in each of them. This leads to the following list of questions:

- What types of waves are transmitted by the different kind of the magnetic structures of different sizes and magnetic field strength?
- At what height do waves develop shocks? How is the wave energy dissipated into the medium?
- How do the frequency spectra change with height in the magnetic structures of different types? What are the mechanisms of these changes? Can direct evidence for mode conversion be found?
- What is the contribution of different frequency ranges to the chromospheric heating?
- Can the photospheric sources of the chromospheric oscillations be uniquely identified? Can direct links between these two layers be found?
- What are the respective contributions of acoustic and magnetic heating?

In order to find the answers to these questions, it is of crucial importance to observe simultaneously several photospheric and chromospheric spectral lines. The topology, field strength, and magnetic flux of the magnetic fields have to be known in the photosphere and in the chromosphere. The evolution of the fields has to be followed closely, as well as that of the purely thermodynamic quantities like temperature and flows. Regardless of the type of magnetic structure observed, high spatial and temporal resolution, large field of view, and high spectropolarimetric precision are always required.

- How do active regions emerge from the convection zone?

The origin of active regions is thought to be the emergence of toroidal flux tubes initially stored at the bottom of the convection zone, the site of the solar dynamo. After being amplified at the bottom of the convection zone, the toroidal magnetic field becomes unstable and rises buoyantly to the solar surface, where it emerges in the form of bipolar active regions. The growth of active regions is rapid and complex, affecting the photosphere, the chromosphere and the corona. At moderate spatial resolution, one observes the formation of arch filament systems in the chromosphere, strong flows in the photosphere and chromosphere (suggesting the occurrence of convective collapse), and flare activity. So far, the interrelation of these events and the topology of the field during the emergence of active regions have not been well characterised because of three main reasons: (a) no ground-based telescope or spacecraft has the capability to measure magnetic fields in the photosphere and chromosphere simultaneously, with high temporal and angular resolution; (b) the evolution of the region must be followed for at least the first 2-3 days; and (c) quite large fields of view are required to cover the full active region.

Observations with EST can be used to understand the sequence of events during the emergence of active regions at their intrinsic scales, to determine the evolution of the photospheric magnetic field and flows, and to investigate the interaction of the newly emerged flux with the pre-existing chromospheric magnetic field (a potential source of heating in the upper layers via magnetic reconnection). The observations will make it possible to test the predictions of 3D MHD simulations of the eruption of flux tubes from below the photosphere into the upper atmospheric layers.

Some important parameters related to the working of the solar dynamo that should be investigated with EST include: (a) the amount of twist in the flux tubes that rise buoyantly from the convection zone; (b) the onset of fragmentation processes and the occurrence of convective collapse as flux tubes reach the photosphere; (c) the changes in magnetic connectivity taking place in the photosphere and chromosphere during the emergence; and (d) the rate of chromospheric heating due to magnetic reconnection.

To achieve all these aims, and others, EST must combine the best of present European facilities:

- Excellent narrow-band and polarimetric imaging capabilities, such as those of the 1-m Swedish Solar Telescope (SST), by using simple optical design and highly performing adaptive optics.
- Open design, as that of the Dutch Open Telescope (DOT), to exploit the favourable winds at the Canary Islands.

- Robust and user-friendly adaptive optics systems, such as the one operating at the German VTT .
- Multi-line spectroscopy, such as that operated by THÉMIS. .
- Visible and near-infrared simultaneous spectropolarimetry, such as the combination of Visible Polarimeters (VIP)-Tenerife Infrared Polarimeter (TIP) and Polarimetric Littrow Spectrograph (POLIS)-TIP at the VTT.
- Efficient narrow-band tunable filters, such as the Interferometric Bidimensional Spectrometer (IBIS) or the Triple Etalon Solar Spectrometer (TESOS).
- Simultaneous control of polarimetric imaging and spectrograph instruments, such as in the VTT.
- Post-processing techniques to compensate for residual seeing effects in multi-wavelength polarimetric images, such as developed for the SST.

EST is expected to improve by a considerable factor the presently achieved spatial resolution. The operation of several narrow-band tunable visible and near-infrared imaging instruments together with grating spectrographs, all with polarimetric capabilities, will make EST a unique telescope for simultaneous observations of photospheric and chromospheric layers and for studies of the time evolution of the three-dimensional structure of solar magnetic fields. Its performance will only be comparable to that of the American ATST. Solar studies will benefit tremendously from having two telescopes with a similar power capable of giving an almost full temporal coverage of solar phenomena. Unlike ATST, EST will not have special requirements to observe the solar corona, making a simpler optomechanical design feasible, with an on-axis structure and less stringent stray light constraints. With this solution, the polarimetric calibration of the telescope, which is fundamental for an accurate determination of the magnetic field vector, will be easier than for off-axis telescopes. A single polarimetric unit may be located in the primary focus. This polarimetric unit would be shared by all instruments (or those selected by the observer). Another option is that each instrument may have its own polarimeter. In this case a unit for polarimetric calibration may be placed in the primary focus. Again this is a flexible setup and observers may decide the optimum configuration for their interests.

Flexible setups will allow the observer to use multiple configurations with minimal optical changes. The observer may decide to use many instruments simultaneously, with broad-band beam splitters for sharing light between the instruments, or, choose to use dichroic beam splitters to distribute all light of a specific wavelength to a specific instrument, thereby maximising the photon flux for each wavelength.

To maximise the efficiency, the optical design of the telescope must integrate in a natural way all the active and adaptive optics, minimising the number of optical

surfaces. This optimisation comes with two advantages. On the one hand, the total throughput of the system and photon transmission will be maximised. On the other hand, wavefront distortions introduced by the optical surfaces will be kept to a minimum. A superb image quality must be one of the major strengths of EST.

Future state-of-the-art ground-based telescopes that will be operating in the coming decades (e.g. GREGOR or ATST) share some of the science goals of EST. EST will have the capability to operate several of its focal-plane instruments simultaneously, and thus allow to study the magnetic connection of the solar magnetic field throughout the solar atmosphere and the interaction of the magnetic field with the moving plasma. This will ensure that these facilities are complementary, resulting in greater scientific return.

Space missions are usually intended to observe wavelengths not accessible from the ground (such as X-rays or the far ultraviolet). Our knowledge of the corona has considerably increased as a result of these instruments. Nonetheless, even if the magnetic connectivity between the different layers of the solar atmosphere is the primary goal of most of them, the study of the deep atmosphere is generally limited to photospheric longitudinal or vector magnetograms (difficult to calibrate accurately in terms of magnetic field) with moderate spatial resolution. The recently launched Hinode mission, with its 50-cm optical telescope, is equipped with a visible spectropolarimeter with a spatial resolution of 0.3 arcsec, to study the photospheric magnetism. The 1-metre balloon-borne solar telescope SUNRISE, will measure the photospheric and the chromospheric magnetic field, but only during its 10 day mission in 2009. There are no currently planned missions that will achieve the resolution of EST in this area during the next decade. To achieve a complete understanding of solar magnetism and its effects in the heliosphere it is necessary to make the connection between all the layers of the atmosphere by combining ground-based and space-based observations.

## 5 Technical challenges for EST

From a technical point of view, a 4-metre class solar telescope is a challenge. Despite all the knowledge and expertise in building and operating solar telescopes that has been acquired in Europe during the last decades, the construction of a large facility is not without risks, and a substantial effort must be made to make it a reality. Presently, there are no operational telescopes of this category in the world.

Thermal effects due to solar radiation are tremendous. To minimise the wavefront distortion in the telescope itself and its surroundings, a careful thermal control of the environment of the telescope is critical. Mirror seeing, turbulence induced by the telescope structure and the floor on which it stands, and wind buffeting on the mirrors and the structure (causing spatial and temporal vibrations) are, among other

problems, to be studied together as a single system to get a coherent design. All these issues are directly related to the existence or not of a dome protecting the telescope during diurnal operation.

Analyses have demonstrated that mirror surfaces need to be cooled to a temperature very close ( $\sim 0.3$  degrees) to ambient air temperature for an efficient reduction of turbulence close to them. This cooling may need to be combined with an additional, natural or forced, air flow to sweep any residual turbulent cells. An open telescope can take advantage of natural winds better than a dome-based solution, even if this has flaps with controlled openings. Since the ambient temperature varies considerably along the day, the mirrors have to be constantly tuned to maintain the required temperature difference. Also, there always exists a time lag between the moment when the cooling system is set to a given temperature and the moment when the mirrors reach it. The inertia of the mirrors to temperature changes has to be taken into account, so that at every instant the temperature difference is always kept within specifications. Mirror thicknesses and materials are thus closely tied-up to their corresponding cooling mechanisms.

The heat load on the primary mirror, M1, is determined by the solar radiation and can consequently be analysed independently of the rest of the system. This is not the case for the rest of the optical surfaces. The amount of radiation they receive depends directly on the field-of-view that is transmitted by the heat rejector. A trade-off must be reached to have the largest possible field of view, without compromising the scientific objectives and the technical solutions for an efficient cooling.

The deformation of the figure of the primary mirror is also a subject that requires a close analysis. Since it is expected that it will be a light mirror, to reduce the overall mass of the telescope and to reduce the heat capacity of the mirror, it will tend to change its shape under conditions of varying gravity vector, as the telescope changes elevation during a diurnal observation. This deformation can be measured and corrected by means of a look-up table. More critically, its low weight and large surface will make it very sensitive to deformations due to wind effects. The buffeting will produce a vibrational spatio-temporal spectrum that will depend on the material and thickness of the mirror. An active compensation of M1 will be required to maintain its figure, using an accurate wavefront sensor. The optimum location of the actuators and their specifications will be obtained after a careful finite-element analysis. To avoid any mechanical or thermal incompatibility between the support of M1 and its cooling mechanism, they will need to be designed together.

The deformation of other mirrors (especially that of the secondary, M2, or of the tertiary, M3, which deviates the beam along the elevation axis) cannot be ignored, but it is substantially less critical, since it depends on the mirror aspect ratio. However, thermal conduction is related to their absolute thickness. M2 and M3 can thus be relatively thicker thereby reducing deformation while having tolerable thermal properties.

The telescope structure is also affected by thermal effects and wind buffeting. Direct sunlight will make it suffer expansions and contractions throughout the day that will modify the separation between the primary and secondary mirrors. Differential heating, or slight inhomogeneities, of the structure will also produce changes in the relative orientation between them. An active M2, built on a hexapod, is necessary to keep fixed its distance and orientation with respect to M1.

Wind has two effects on the telescope structure. On the one hand, it will excite its natural eigenfrequencies. The active supports of M1 and M2 will be in charge of the compensation of these vibrations. A careful control of the spatio-temporal variations will be required, to determine which actuators are to be applied. On the other hand, a laminar flow crossing the telescope will be deformed by the presence of the bars of the structure, giving rise to a turbulent flow. Temperature fluctuations associated with this flow should be kept to a minimum, since they are the responsible for refractive index and optical length variations, i.e., for the distortion of the wavefront.

The spider supporting M2 also needs special consideration. It needs to be stiff enough to minimise its deformation but must have a small cross-section because its shadow will be seen by the subaperture images generated with the adaptive optics system. An alt-azimuthal structure is expected for EST, due to its large size and weight, and will give rise to a rotation of the pupil in a fixed reference frame, such as the location of the wavefront sensor. The spider projection on each subpupil of the AO system should represent a small fraction of the subpupil area, so that they are all useful at every instant. The cooling system of M2 will be guided through the telescope structure and, in particular, through the spider. A large flux of the cooling fluid implies a large section of the spider, with a negative impact on the AO performance. Since the field of view determines how much energy needs to be removed from M2 to keep it at the desired temperature, a close relation between the designs of the heat rejector, the spider and AO system is mandatory.

Following all the above reasoning, telescope structure (including the spider for M2), the active system of M1 and M2, the mirror cooling mechanisms, the heat rejector and the AO system are part of the same system and must be designed very tightly to match each other and minimise the impact of heat and wind loads.

There is another factor that critically affects the mentioned subsystems: the dome. The DOT has demonstrated that an open telescope can operate very efficiently and, with the help of natural wind, diffraction-limited images can be obtained. It is not clear whether this experience can be extrapolated to a 4-metre class telescope and an in-depth analysis needs to be performed. In an open telescope, the whole telescope (structure and mirrors) is affected by wind. The image-quality performance of the telescope must be guaranteed for all possible wind directions and telescope orientations. In addition, the whole floor where the telescope is located is heated by the solar radiation and must be actively cooled to avoid locally generated seeing. If a dome is installed to protect the telescope from wind, then the problem is transferred

to the local seeing generated by the dome itself. A critical decision about the dome must be made by the project, since directly affects the technical requirements of the whole telescope and AO systems.

The pier and the building have to be designed in such a way that the vibrations of the building or the pier induced by wind are small enough to be corrected by the AO.

As mentioned above, the adaptive optics system is responsible for evaluating the control signals of M1 and M2 and sending them to their actuators in case the mirrors or the structure itself suffer from deformations. The AO must, in addition, evaluate the wavefront distortions induced by the air along the optical path, inside the telescope, in its surroundings and in the free atmosphere. Turbulence has typical variation frequencies of a few hundred Hz, larger than those of mechanical vibrations of solid structures. For this reason, a fast tip-tilt mirror and, at least, one deformable mirror are required in the optical path to compensate for turbulence-induced wavefront inclination and deformation. At visible wavelengths, the isoplanatic angle is expected to be rather small for a 4-metre class telescope (a few arcsec). A Multi-Conjugate AO (MCAO) system with, at least, a second deformable mirror, is required to have a larger diffraction-limited field of view. In the near-infrared, this parameter is expected to be considerable larger. The number of layers of the atmosphere that need to be corrected is at present unknown.

Finally, instruments must be addressed as a whole. These include broad-band imagers, with phase diversity and/or speckle imaging capabilities, in several wavelengths. Etalon-based tunable imagers, covering from short visible wavelengths to the near-infrared, seem to be the most promising candidates for narrow band imaging. Independently of whether a collimated or a telecentric solution is adopted, the selected field of view will be critical. In addition, one or several long-slit or integral-field spectrographs must also be included. Narrow-band imagers usually give much better imaging quality than spectrographs, but these give more coherent spectral information. For this reason, the combination of narrow band imaging filters and spectrographs will be fundamental for many observing programs. The telescope optical design must give the capability of using all these instruments simultaneously or only a few of them, with an optimised light distribution.

*Acknowledgements:* This report is the consequence of the work of many people, whose names can not be explicitly listed without surpassing the space reserved to the author's list. Spanish solar astronomers first contributed to the document "Spanish Science Vision for EST: Magnetic coupling of the solar atmosphere". Later, many European solar astronomers contributed to complement and improve it for the presentation of the EST project to the EC under the FP-7 programme. This report is based on these two documents.



# ROBOTIC ASTRONOMY AND ITS APPLICATION TO THE STUDY OF GAMMA-RAY BURSTS

ALBERTO J. CASTRO-TIRADO

*Instituto de Astrofísica de Andalucía-Consejo Superior de Investigaciones  
Científicas (IAA-CSIC), E-18008 Granada, SPAIN*

**Abstract:** An overview of Robotic Astronomical facilities (especially in Spain) is presented. The study focuses on two aspects: the control software (one of such example being the RTS2 system) and the network of BOOTES robotic telescopes, partly devoted to the study of gamma-ray burst counterparts at optical and near-infrared wavelengths. This potential application of small/medium size robotic telescopes will shed light on the high redshift Universe and should be used for triggering larger size instruments in order to perform more detailed studies of host galaxies and intervening material on the line of sight.

**Keywords:** Robotic Astronomy – Control Software – Gamma-ray Bursts.

## 1 Introduction

Robotic astronomical observatories (RAOs hereafter) were first developed in the 1990s by astronomers after electromechanical interfaces to computers became common at observatories. Following [1], let us introduce some definitions first:

- **Robot:** A mechanical system which executes repetitive tasks with good accuracy with human assistance. Example: Industrial robotic arm.
- **Teleoperated Robot:** A mechanical system which executes a given task with good accuracy and that can be modified with human assistance. Example: Submarine research robots.
- **Intelligent Robot:** A mechanical system which executes a task with good accuracy and is able to adapt itself to changes during the task execution without any kind of human assistance. Example: Rovers devoted to planetary research.

## 2 Robotic Astronomical Observatories: a brief history

The 1985 book *Microcomputer Control of Telescopes* by R. M. Genet and M. Trueblood [2], was a landmark engineering study in the field. Since the commissioning of the Bradford telescope (in 1993) [3] and the Iowa Telescope (in 1997) [4], many researches and companies have put considerable effort in making robust systems.

The *first robots* were the telescopes with an absolute positioning control and guiding systems, and the automatic weather stations, introduced in astronomical observatories.

The *first robotic astronomical observatories* were those ones which were able to integrate and coordinate the different automatic subsystems at the observatory (telescope, dome, weather stations). But they require human assistance (teleoperation) for the taking of decisions regarding a given task and/or its supervision.

The *intelligent robotic astronomical observatories* are the following step, where human assistance in the taking of decisions is replaced by an artificial intelligent system. These are being developed nowadays.

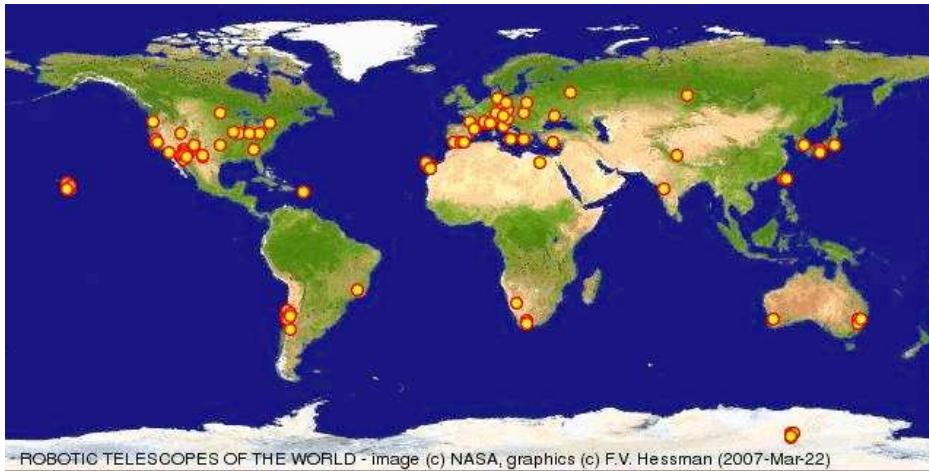


Figure 1: The RAOs location in the world. Adapted from Hessman [6].

### 3 RAOs worldwide

Based on the compilation collected by F. V. Hessman [6], there are about 100 RAOs worldwide (see Fig. 1), with 35 of them being located in Europe. Some examples are:

- ROTSE (UM & LANL, USA): A network of four 0.45 m diameter telescopes around the world, devoted to the search for optical transients [7].
- RAPTOR (LANL, USA): An array of telescopes that continuously monitor about 1500 square degrees of the sky for transients down to about 12th magnitude in 60 seconds and a central fovea telescope that can reach 16th magnitude in 60 seconds. Search for optical transients (OTs). See Fig. 2 [8].
- REM (Italy): Is a rapid reaction near-infrared (nIR) robotic telescope [9] dedicated to monitor the prompt afterglow of Gamma Ray Burst (GRBs) events [10].
- PAIRITEL (SAO, USA): It is a 1.3m telescope devoted to the study of nIR transients [11] by means of simultaneous JHK imaging [12].



Figure 2: The RAPTOR wide-field telescopes system [6].

- ROBONET (participated by 10 UK Universities) is a network of three 2m class robotic telescopes (see Fig. 3). The main aims are to detect cool extra-solar planets by optimised robotic monitoring of Galactic microlens events. In particular, to explore the use of this technique to search for other Earth-like planets. Another goal is to perform detailed studies of GRBs [13].

## RoboNet-1.0



Figure 3: The 2m RoboNet Network. Adapted from [13].

## 4 RAOs in Spain

Amongst the  $\sim 35$  RAOs existing in Europe (see Fig. 4), a dozen of them are located in Spain, with some of them being automated systems and few others being robotic ones.

The Spanish automated systems are the Carlsberg telescope (since 1983), the IAA Tetrascop (4 x 0.35m) at OSN (2001-05) and La Sagra (since 2006), the 0.45m Astrograph at La Sagra (since 2007) and the DIMMA (IAC), an automated seeing monitor in operation since 2007.

The Spanish robotic systems are the 0.2m and 0.3m BOOTES-1 telescopes (since 1998), the 0.3m BOOTES-2 telescope (since 2001), the 0.6m BOOTES-IR telescope (since 2004), the 0.6m TROBAR telescope (since 2004), the 0.8m MONTSEC telescope (since 2005) and the 0.4m, 0.5m and 0.5m belonging to the CAB Robotic Telescope Network (since 2004).

We provide additional details for some of them:

- The Circulo Meridiano Carlsberg was initiated by KUO, IoA and ROA, with ROA being the only institution that run the instrument nowadays. It is an



Figure 4: The RAOs location in Europe. Adapted from [6].

automated telescope placed at La Palma (Canary islands), which had first light in 1983 (see Fig. 5). It allows to observe between 100,000 and 200,000 stars a night, down to  $r'=17$ . This will give accurate positions of stars, allowing a reliable link to be made between the bright stars measured by *Hipparcos* and the fainter stars seen on photographic plates (as measured by the APM and similar measuring machines). The current area of the survey is between  $-30$  and  $+50$  degrees in declination and is completed [14].

- TROBAR (UV) is a 0.6m diameter robotic telescope located in Aras del Olmo (Valencia), which had first light in 2004. It is devoted to astroseismology and extrasolar planet research. NEOs and GRBs studies are also part of the scientific programme [15].
- The largest diameter robotic telescope in continental Spain is the 0.8m diameter telescope at Observatori Astronomic del Montsec (participated by UB, UPC, CSIC, Consorci del Montsec and Fundació Joan Oró). It had first light in 2005 and makes use of the TALON control system [16].



Figure 5: The Carlsberg meridian Circle, at La Palma, is operated by the Real Observatorio de la Armada in San Fernando (Cádiz).

- The CAB/INTA/CSIC Robotic Telescope Network is formed by a 0.4m diameter telescope in Torrejón de Ardoz (Madrid), a 0.5m diameter telescope in Calatayud (Zaragoza) and a 0.5m diameter robotic telescope in Calar Alto (Almería). [17].
- BOOTES-1 and BOOTES-2 (participated by INTA/CSIC/AUS/CVUT) started with robotic 0.3m and 0.2m diameter telescopes and wide-field lens systems, having first light in Huelva (1998) and Málaga (2001) respectively. The telescopes will be upgraded to 0.6m telescopes in 2008-2009.
- BOOTES-IR/T60 (CSIC) is a robotic 0.6m diameter telescope at Observatorio de Sierra Nevada (Granada), which had first optical light in 2004 and first near-infrared (nIR) light in 2007. Simultaneous optical/nIR imaging is foreseen for late 2008. Additional details for both BOOTES and BOOTES-IR are given below.

#### 4.1 BOOTES

BOOTES, the **B**urst **O**bserver and **O**ptical **T**ransient **E**xploring **S**ystem, is mostly a Spanish–Czech international collaboration that works to fill in the space that actually exists in rapid variability Astronomy. It is specially aimed towards the detection and study of the optical transients that are generated in conjunction with the elusive



Figure 6: The BOOTES-1 0.2m and 0.3m diameter telescopes at Instituto Nacional de Técnica Aeroespacial in Mazagón (Huelva). They will be replaced by a 0.6m telescope.

GRBs. It saw first light in 1998 [18] being one of the pioneering robotic observatories for OT follow ups [19]. There are two 250 km distant BOOTES stations. Thus, using parallax, it can discriminate against near Earth detected sources up to a distance of  $10^6$  km.

BOOTES-1 in Mazagón (Huelva) has two domes (1A and 1B), three Schmidt-Cassegrain telescopes (Fig. 6) and several wide field cameras. Following complementing schemes, all instruments carry out systematic explorations of the sky each night.

BOOTES-2, located near Málaga is in operation since 2002. It has one 30cm telescope with an attached wide-field camera. The station may observe in standalone as well as in parallel stereoscopic modes together with BOOTES-1. An ultra-light weight 0.6m telescope (TELMA) is replacing the existing one in Spring 2008 [20]. See Fig. 7.

Both stations are operated under the RTS2/Linux control system (see section 6).

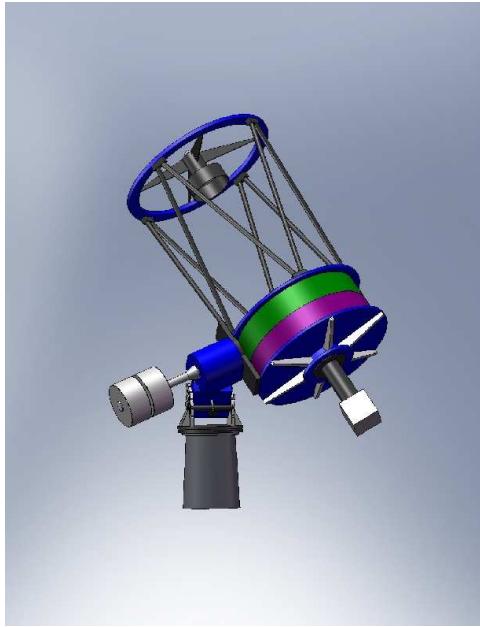


Figure 7: The TELMA ultra-light weight telescope concept for the BOOTES-2 station in Málaga (Spain).

## 4.2 BOOTES-IR

BOOTES-IR, the **B**urst **O**bserver and **O**ptical **T**ransient **E**xploring **S**ystem in the near-**I**nfra**R**ed, is the extension of the BOOTES project towards near-IR wavelengths thanks to a nIR camera developed in the context of Spain's Programa Nacional de Astronomía y Astrofísica, placed in 2006 at the 60 cm telescope at the Observatorio de Sierra Nevada, under a controlled dome, also developed in the context of the Project (see Fig. 8).

BOOTES-IR was first proposed in 2001. The enclosure was built atop Sierra Nevada in the Summer of 2003. The telescope was installed at the end of 2004 and first (optical) light was obtained in 2005. Since then the telescope is in commissioning phase and operating with an optical camera, and responding to some alerts within 20-30 s after occurrence. The nIR camera has had first light in 2007 [21].

Thus, BOOTES-IR will be the third astronomical nIR RAO of this kind [22], following REM (opt/nIR) at ESO La Silla Observatory in Chile and PAIRITEL (nIR), but extending its wavelength coverage in the blue optical range.



Figure 8: The BOOTES-IR camera (BIRCAM) attached to the 0.6m robotic BOOTES-IR telescope at the Observatorio de Sierra Nevada.

## 5 Technology with RAOs

### 5.1 Range of apertures

According to recent statistical studies [6] and once instruments planned by 2010 are considered, nearly 50% of RAOs have diameter smaller than 0.25m, while 10% have diameter larger than 1.25m. Nearly 95% are equipped with optical instrumentation, with the remaining fraction being devoted to nIR studies.

### 5.2 Telescope Control Operating Systems and Observatory Managers

*Control Operating Systems* can be divided into commercial or specific ones, which can be open or closed source. For instance, a commercial automatization systems is TCS, developed by Optical Mechanics (OMI), for operating telescopes with diameters in the range 0.4 to 1.0 m [23]. A specific control system is the one built for 10 m Spanish GTC telescope.

Amongst *Observatory Managers* some examples are:

- AUDELA: Developed by A. Klotz et al. (Toulouse), starting in 1995. Open source code. Linux/Windows [24].

- ASCOM: Designed in 1998, by B. Denny (USA), as an interface standard for astronomical equipment, based on MS's Component Object Model, which he called the Astronomy Common Object Model. Mostly used by amateur astronomers, has been also used by professionals, under the Windows operating system. It is widely used in supernovae and minor planet searches [25].
- RTS2: The Robotic Telescope System version 2, is being developed by P. Kubánek, (Ondrejov/Granada) starting in 2000. The source code is open. It works under Linux/Windows (command line and graphical interface foreseen). Widely used in GRB searches.
- INDI: The Instrument Neutral Distributed Interface (INDI) was started in 2003. In comparison to the Microsoft Windows centric ASCOM standard, INDI is a platform independent protocol developed by E. C. Downey (USA). The source code is open too. Not so widely spread as the upper layer interface was not done [26].

Observatory Managers can also work as open or close loop systems. In an open loop system, a robotic telescope system points itself and collects its data without inspecting the results of its operations to ensure it is operating properly. An open loop telescope is sometimes said to be operating on faith, in that if something goes wrong, there is no way for the control system to detect it and compensate. A closed loop system has the capability to evaluate its operations through redundant inputs to detect errors. A common such input would be position encoders on the telescope's axes of motion, or the capability of evaluating the system's images to ensure it was pointed at the correct field of view when they were exposed [27].

## 6 RTS2

RTS is a system for complete observatory control. It can be regarded as a turnkey system. Once installed on any telescope, it should run and provide results. RTS consists of three major layers – device, service and monitoring, with components communicating over TCP using a simple text protocol. Detailed design and development history of RTS can be found in [28]. See Fig. 9.

### 6.1 Current RTS2 system operation

The current network of RTS controlled observatories operates as a set of separate nodes. The nodes run every night, paying attention to local conditions via sensors and controlling all aspects of the telescope operations.





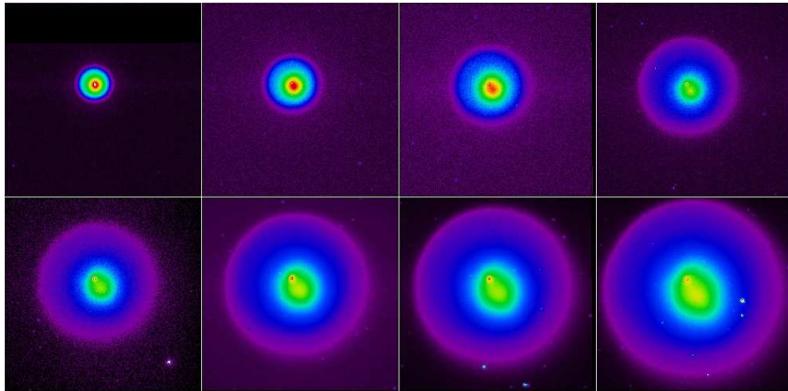


Figure 11: The evolution of comet 17P/Holmes following the October 2007 outburst, imaged on a nightly basis with the 0.3m BOOTES-2 telescope in Spain. The depicted field of view is the same ( $\sim 10' \times 10'$ ) in all frames.

objects in binary systems and blazars should be easily observable. The possibility of simultaneous optical/nIR monitoring will allow us to study in great detail the underlying physical emission mechanisms.

Particularly, RAOs can be used to study comets and asteroids. In the wavelength range of the detector is feasible to analyze the radiative properties of the dust in cometary comae. Especially important can be the physical characterization of Near Earth Objects (NEOs). On the other hand, optical-nIR observations can help to determine the luminous efficiency of meteoritic impact flares on the Moon.

## 8 Study of GRBs

### 8.1 GRBs at the edge of the observable Universe

Cosmology is starting to constrain the nature of the earliest galaxies formed in the Universe but direct observations of galaxies at  $z > 4$  remains largely challenging due to the observational constraints imposed for their large luminosity distances. Long-duration Gamma-ray bursts (long-GRBs), those ones usually lasting more than 5s and originated at cosmological distances, can be used as beacons to point to the location of these high- $z$  galaxies, thanks to their extreme luminosities (with energy releases of  $10^{51}$ – $10^{53}$  ergs).

As the central engines that power these extraordinary events are now considered to be the collapse of massive stars ([35, 36]), they can be used as tracers of star formation.

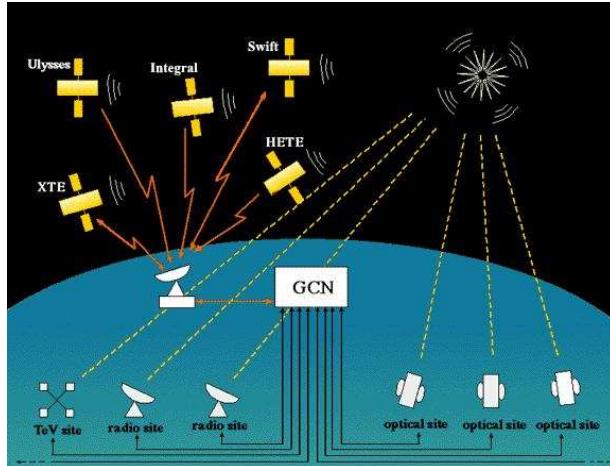


Figure 12: The GCN/Bacodine alert system assures rapid dissemination of GRB alerts thru Internet, so RAOs can quickly respond to them [33].

Early observations of the afterglows (the multiwavelength emission that follows the gamma-ray emission) allow to provide redshifts and additional spectral observation which cannot be derived from the direct observations of the galaxies themselves. This is most essential as some of these host galaxies (the ones a  $z > 6$ ) are responsible of a significant proportion of ionizing radiation during that reionization era.

*INTEGRAL* and *Swift* are providing around 100 detections/yr, whose alerts are promptly distributed by the GCN/Bacodine Network (see Fig. 12). However, none of these missions carry instrumentation devoted to the nIR, that could complete the observations at longer wavelengths. Following the detection of a bright, prompt optical flash for GRB 990123 with  $M_V = -36$  [37], such events were also expected to occur at nIR wavelengths, as was proven for GRB 041219 [38].

But since the first GRB with a fading X-ray afterglow *without* an optical counterpart was detected by *BeppoSAX* in Jan 1997, the number of similar events has increased significantly since the launch of *Swift* in Nov 2004. Nowadays, these dark GRBs seem to constitute a significant fraction ( $\sim 50\%$ ) of the GRB population (e.g. [39]).

This latter population of dark GRB should contribute significantly to the hidden star formation rate in the Universe [40]. Important scientific goals for some RAOs are the study of dark GRBs and the “hunt” for ultra-high  $z$  events.

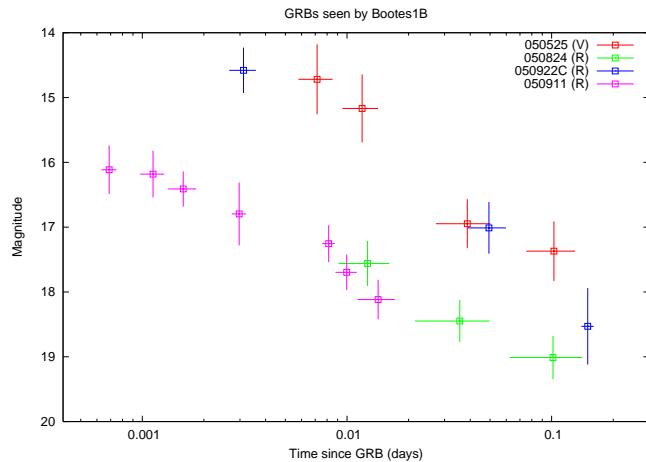


Figure 13: Optical afterglows lightcurves of some GRBs detected by BOOTES and rapidly imaged ( $\sim 1$  min) after the detection by scientific satellites.

## 8.2 Dark GRBs

About 50% of events are not detected in the optical in spite of deep observations being performed minutes/hours after the event. This can be partly explained if the GRB do occur in a high density region in the host galaxy which will extinct the optical emission. Thus, even with no optical afterglow being observed, a bright nIR transient might be recorded and, together with the derived upper limits in the optical, might allow to determine the intrinsic extinction. In fact, it is expected that most of the dust will be sublimated by the prompt UV/optical emission, i.e., it is foreseen that the nIR flash should be observed prior to the optical one, allowing to determine an upper limit to the amount of dust in the surroundings of the GRB progenitor.

## 8.3 Ultra-high $z$ GRBs

There will be a small fraction of events not detected in the optical due to a ultra-high redshift (with  $z > 6$ ), i.e., with the  $\text{Ly}\alpha$  break in the I band ( $0.9 \mu\text{m}$ ) at that particular redshift. As nIR RAOs will cover a redshift range  $6 \leq z \leq 17$  (see Fig. 14), part of these ultra-high  $z$  population should be unveiled if prompt observations are conducted soon after the events. As it has been already pointed out, GRB are a powerful tool for the study of the high  $z$  Universe. The importance of prompt observations is based on the fact that the photons arrival time should be divided by a  $(1+z)$  factor. That

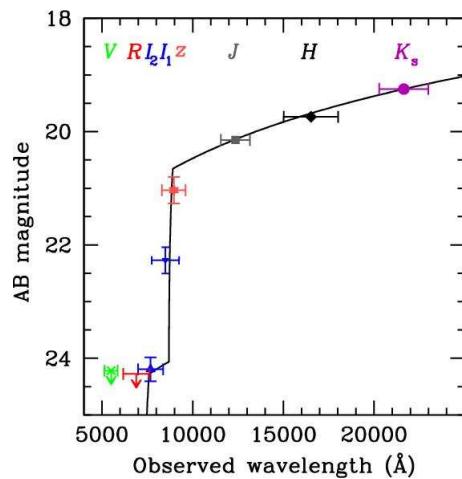


Figure 14: GRB 050904 holds the highest redshift recorded so far for a GRB ( $z = 6.3$ ). The Lyman- $\alpha$  break is quite noticeable in the broadband optical-nIR photometry [41].

is, that the elapsed time since the onset of the event has to be accounted for in the framework of the source. Thus, a GRB at  $z = 4$  being observed 10 min after the event means that it is being detected  $10/(1+z) = 2$  min after the explosion, when the source is extremely bright. Thus, we can say that  $z$  might be an advantage, with a favourable K correction [42].

For instance, a combination of the BOOTES (optical) and BOOTES-IR (nIR) datasets will allow us to distinguish a high  $z$  event. The identification of candidates in a color-color diagram [43] will allow us to discern the most interesting candidates allowing larger size instruments to point to the GRB afterglow while it is still bright enough to ease spectroscopic observations. This will allow to study the distribution of Ly $\alpha$  clouds in the intergalactic medium as function of  $z$ , the metallicity, the interstellar medium in the host galaxy and the intergalactic medium reionization, expected in the  $6 \leq z \leq 17$  range.

#### 8.4 Coordinate observation of GRBs in different filters

Only a few GRBs have exceptionally bright optical counterparts. Observers are of course interested in collecting as much data as possible, with the best possible resolution.

One of the goals of the observers is to take spectra of the transient while it is bright enough, so the transient redshift and other properties can be measured. Using data

taken with different filters, one can construct a spectral energy distribution of the event and estimate object redshift. The networked RTS telescopes (like BOOTES) at favourable locations can simultaneously observe objects in different filters.

The idea is to enable those telescopes to communicate with each other and provide simultaneous images in two or more filters. This system should balance the need to take some data with possibility to take data in multiple filters. It can be achieved by sending commands for taking images in different filters when the system knows that it has at least some images of the event. This kind of decision is best done in a single component – observation coordinator.

The coordinator will be connected to two or more telescope nodes. It will collect information from GCN and from all connected nodes. A node will report to the coordinator when it receives a GCN notice, when it starts its observation and as soon as it gets image passed through astrometry and it contains whole error area of the GRB. It will also report when the transient detection software identifies a possible optical transient.

When the coordinator receives messages about correct observation by two telescopes, it will decide which filter should be followed at which telescope, and send out commands to carry out further observations. The coordinator will periodically revisit its observing policy, and send out commands to change filter accordingly.

As the system is “running against the clock” for the first few minutes after the GRB event, trying to capture the most interesting part of the transient light curve, it cannot wait for the completion of the transient source analysis. In the case of two telescopes, the coordinator will command different filters as soon as it knows that both telescopes have acquired the relevant field. The current astrometry routines take a few seconds to run, and it is expected that observations with different filters can already have started within this time-frame.

## 9 Coordinated World-wide observation of Targets

One step further from the GRB observation is coordinated observation of targets – e.g. observation of variable stars for more than 12 hours (i.e. taking advantage of telescopes in different time zones). The observer should contact the coordinator, and either add a new target, or select a predefined target which he/she wants to observe. The coordinator should list to the observer telescopes which can observe target of his/her choice, and propose filters and exposure times.

The observer could then decide which telescopes to be used, and the coordinator would send observation requests to the nodes, and collects back information about observation progress. Currently only observer-selected coordinated observations are envisioned. Once that will work properly, observer can be replaced by a network scheduling software.

## 10 Conclusions

Robotic Telescopes are opening a new field in Astrophysics in terms of optimizing the observing time, with some of them being able to provide pre-reduced data. The big advantage is that they can be placed in remote locations where human life conditions will be hostile (Antartica now, the Moon in the near future).

Technological development in different fields is pretty much involved and some of the robotic astronomical observatories are moving towards intelligent robotic astronomical observatories.

One immediate application of small/medium size robotic telescopes is the study of GRBs, which can be considered the most energetic phenomenon in the Universe. In combination with space missions like *Integral* and *Swift* (nowadays), *GLAST* (to be launched in mid 2008) and *EXIST* (if approved) should be used for triggering larger size instruments in order to perform more detailed studies of host galaxies and intervening material on the line of sight. These RAOs will provide a unique opportunity to unveil the high- $z$  Universe in the years to come.

*Acknowledgements:* I thank P. Kubánek and the entire ARAE team at IAA-CSIC for very fruitful conversations. This work is supported by the Spanish MCyT AYA 2007-63677 and Junta de Andalucía P06-FQM-02192 and P07-TIC-03094.

## References

- [1] Berná Galiano, J.A. and Cotillas López, A. 2004, “Astrofísica Robótica en España”, Edited by A.J. Castro-Tirado, B.A. de la Morena and J. Torres, 51
- [2] Trueblood, M. and Genet, R.M. 1997, in “Telescope control” (2nd edition), Willmann-Bell
- [3] McKinnon, D.H. and Mainwaring, A. 1993, “The Charles Sturt University Remote Telescope Project: Astronomy for Primary School Students” <http://www.atnf.csiro.au/pasa/17.2/mckinnon/>
- [4] <http://phobos.physics.uiowa.edu>
- [5] Barthelmy, S.D. 2003, AAS 35, 765
- [6] Hessman, F.V. 2007, <http://www.astro.physik.uni-goettingen.de/hessman/>
- [7] <http://www.rotse.net/>
- [8] <http://www.raptor.lanl.gov/>
- [9] Covino, S., Stefanon, M., Sciuto, G. et al. 2004, “Ground-based Instrumentation for Astronomy”, edited by A.F.M. Moorwood and I. Masanori. Proceedings of the SPIE, Vol. 5492, 1613
- [10] <http://golem.merate.mi.astro.it/projects/rem/>
- [11] Bloom, J.S., Starr, D.L., Blake, C.H. et al. 2006, “Astronomical Data Analysis Software and Systems XV”, ASP Conference Series, edited by C. Gabriel, C. Arviset, D. Ponz, and E. Solano, Astronomical Society of the Pacific 351, 751

- [12] <http://astro.berkeley.edu/~jbloom/Autotel/>
- [13] <http://www.astro.livjm.ac.uk/RoboNet/>
- [14] <http://www.ast.cam.ac.uk/~dwe/SRF/camc.html>
- [15] <http://www.uv.es/obsast/es/instrument/trobar.html>
- [16] Fernández, D., Isern, J., Palau, X. and Torra, J. 2004, AN 325, 657
- [17] Galadí, D. and Pinés, M.R. 2004, “Astrofísica Robótica en España”, edited by A.J. Castro-Tirado, B.A. de la Morena and J. Torres, 119
- [18] Castro-Tirado, A.J., Hudec, R. and Soldán, J. 1996, “The Third Hunstville Symposium on Gamma-Ray Bursts”, ed: Kouveliotou, C. et al., AIP Conf. Proc. 384, 814
- [19] Castro-Tirado, A.J., Soldán, J., Bernas, M. et al. 1999, A&AS 138, 583
- [20] <http://www.iaa.es/bootes/>
- [21] <http://www.iaa.es/bootes-ir/>
- [22] Castro-Tirado, A.J., Cunniffe, R., de Ugarte Postigo, A. et al. 2006, “Ground-based and Airborne Telescopes”, edited by Stepp, Larry M.. Proceedings of the SPIE, Volume 6267, 626701
- [23] [http://www.opticalmechanics.com/observatory\\_telescopes/OML\\_telescopes.html](http://www.opticalmechanics.com/observatory_telescopes/OML_telescopes.html)
- [24] [http://www.audela.org/english\\_audela.php](http://www.audela.org/english_audela.php)
- [25] <http://ascom-standards.org/>
- [26] <http://linux.softpedia.com/get/Science-and-Engineering/Astronomy/INDI-13292.shtml>
- [27] Kubánek, P., Jelínek, M., Vítek, S., de Ugarte Postigo, A. et al. 2006, Proceedings of the SPIE 6274, 59
- [28] <http://lascaux.asu.cas.cz/rts2/>
- [29] Barthelmy, S.D. 2003, “American Astronomical Society Meeting Abstracts”, vol. 202
- [30] Jelínek, M., Prouza, M., Kubánek, P. et al. 2006, A&A 454, L119
- [31] de Ugarte Postigo, A., Gorosabel, J., Jelínek, M. et al. 2006, GCN 5290
- [32] de Ugarte Postigo, A., Jelínek, M., Gorosabel, J. et al. 2004, “Astrofísica Robótica en España”, edited by A.J. Castro-Tirado, B.A. de la Morena and J. Torres, 35
- [33] [http://gcn.gsfc.nasa.gov/gcn\\_main.html](http://gcn.gsfc.nasa.gov/gcn_main.html)
- [34] Castro-Tirado, A.J., Mateo Sanguino, T.J., de Ugarte Postigo, A. et al. 2004, “5th INTEGRAL Workshop on the INTEGRAL Universe”, ESA SP-552, 637
- [35] Hjorth, J., Sollerman, J., Moller, P. et al. 2003, Nature 423, 847
- [36] Campana, S., Romano, P., Covino, S. et al. 2006, A&A 449, 61
- [37] Akerlof, C., Balsano, R., Barthelmy, S. et al. 1999, Nature 398, 400
- [38] Blake, C.H., Bloom, J.S., Starr, D.L. et al. 2005, Nature 435, 181
- [39] Filliatre, P., Covino, S., D’Avanzo, P. et al. 2006, A&A 448, 971
- [40] Totani, T. 1999, ApJ 511, 41
- [41] Tagliaferri, G., Antonelli, L.A., Chincarini, G. et al. 2005, A&A 443, L1
- [42] Lamb, D.Q. and Reichart, D.E. 2000, ApJ 536, L1
- [43] Gorosabel, J., Fynbo, J.U., Hjorth, J. et al. 2002, A&A 384, 11



# THE MARTIAN UPPER ATMOSPHERE

FRANCISCO GONZÁLEZ-GALINDO<sup>1,2</sup>, FRANCOIS FORGET<sup>1</sup>, MÓNICA ANGELATS I COLL<sup>1</sup>

*1 Laboratoire de Météorologie Dynamique, CNRS, Paris, France*

MIGUEL ÁNGEL LÓPEZ-VALVERDE<sup>2</sup>

*2 Instituto de Astrofísica de Andalucía-Consejo Superior de Investigaciones Científicas, E-18008 Granada, SPAIN*

## **Abstract:**

The most relevant aspects of the Martian atmosphere are presented in this paper, focusing on the almost unexplored upper atmosphere. We summarize the most recent observations concerning this region, as well as the numerical models used to its study.

Special attention is devoted to the only ground-to-exosphere General Circulation Model existing today for Mars, the LMD-MGCM. The model and its extension to the thermosphere are described and the strategies used for its validation are shortly discussed. Finally, we briefly present some comparisons between the results of the model and the observations by different spacecrafts.

**Keywords:** Mars atmosphere – Thermosphere – General Circulation Models.

## **1 Introduction**

### **1.1 Main features of the Martian atmosphere**

Our knowledge of the Martian atmosphere has increased dramatically in the last 3-4 decades as a result of an international effort of exploration (more than 20 spacecrafts have been launched to Mars since the first secret attempts by the Soviet Union in 1960, although only about half of them have been successful) and of the increasing sophistication of the theoretical models devoted to its study. Nowadays we know that the Martian atmosphere, mostly composed of CO<sub>2</sub>, is very thin (pressure surface of about 6 mb) when compared to the Earth (1 bar) and Venus (95 bar) [1] and has a surface temperature that oscillates between 140 and 300 K [2]. The inclination of its axis of rotation (similar to the terrestrial) and the high eccentricity of the Martian orbit induce a seasonal cycle more intense than the terrestrial one.

A typical thermal profile for the Martian atmosphere is shown in Figure 1. Due to the small amount of water in the Martian atmosphere, the decrease of temperature with altitude in the troposphere follows the dry adiabatic [3]. The Martian

atmosphere does not have an stratosphere like the terrestrial one. The reason is that the amount of ozone in the Martian atmosphere is too small to produce a noticeable heating. However, when the dust load is high enough, the heating induced by the dust can produce thermal inversions similar to a stratosphere. The lower thermosphere is characterized by a strong increase of temperature with altitude, due to the absorption of UV solar radiation, while in the upper thermosphere the temperature tends to an asymptotic value due to the high efficiency of the thermal conduction, that suppresses temperature gradients.

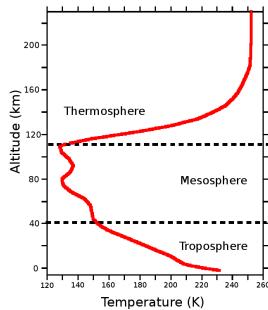


Figure 1: Typical temperature profile of the Martian atmosphere

There are two features of the Martian atmosphere that make it unique in the solar system: the CO<sub>2</sub> cycle and the dust storms.

The CO<sub>2</sub> cycle consists in the interchange of this constituent between the atmosphere and the polar reservoirs in response to the annual change in the insolation of the surface [4]. In winter, the atmospheric temperatures in the polar regions are so low that CO<sub>2</sub> condenses and is deposited on the surface, subliming in spring with the rising temperatures. This cycle affects an important fraction (about one third) of the atmospheric CO<sub>2</sub>, producing remarkable seasonal variations in the surface pressure, that were already detected by the Viking landers [5].

The importance of the presence of dust in the Martian atmosphere was first unveiled by the soviet Mars 2 and Mars 3 spacecrafts, lost due to a dust storm [6]. Today we know that the amount of dust suspended in the atmosphere is higher during Southern hemisphere summer (corresponding to the perihelium of the Martian orbit), with an annual cycle approximately repetitive. However, important interannual variations can occur [7]. During perihelium local dust storms are formed, that can eventually join and produce a global, planet-encircling dust storm. The mechanisms to form these global storms are currently not well known. The dust absorbs the infrared radiation and can affect the thermal structure of the atmosphere [3].

## 1.2 The Martian upper atmosphere

On behalf of the intense international exploratory effort from the second half of the 20th century, it is only in the last years that the upper atmosphere has begun to be explored. In this work, we will refer as “upper atmosphere” to the region with altitudes between about 80 and 250 km, that is, referring to the Figure 1, the upper mesosphere and the thermosphere.

During the last decade there has been a growing interest in this region. The reasons are twofold. First, this region is the scenario of important physical, dynamical and chemical processes, as the absorption of ultraviolet (UV) radiation coming from the Sun, that is the primary heating source of the upper atmosphere [8], the photochemistry, that induces the photodissociation of molecules like  $\text{CO}_2$  and  $\text{O}_2$  into simpler molecules and/or atoms, and the escape, that is essential to understand the long-term evolution of the whole atmosphere. The region possesses a complex dynamics, with interactions between waves, both created in-situ and propagating from below, and the mean flow [9]. Second, it is in this particular altitude range where the spacecrafts perform their aerobraking maneuvers, using the friction with the atmosphere to decelerate the spacecraft up to the velocity appropriate for the insertion in the required orbit. A detailed knowledge of the density structure is necessary to minimize the risks of this maneuver. Given the scarcity of data, theoretical models, like General Circulation Models (GCMs) are essential for this task.

The latest observations have shown a strong coupling between the lower and the upper atmosphere [10]. During its aerobraking, MGS has observed a longitudinal variation of the density (at constant local time and altitude) mainly composed of wave numbers 2 and 3 [11]. It has been shown [12, 13] that the origin of this structure is the interaction of the solar illumination with the topography and the non-linear interactions between waves created in-situ and propagating from below. Mars Odyssey has detected, during its aerobraking phase, an increase of temperature with latitude when moving towards the winter pole [14]. The origin of this thermospheric polar warming is a downwelling from the upper thermosphere due to an intense interhemispheric transport, that produces an adiabatic warming [10]. The intensity of this warming is modified by the dust amount in the lower atmosphere [15]. SPICAM on board Mars Express has detected for the first time in Mars the UV emissions of the NO molecule in the nightside [16]. The peak emission is located between 60 and 100 km, with no clear trend with the latitude, longitude, local time or solar activity. This nightglow is produced by the recombination of N and O atoms, that are transported to the nightside mesosphere from the dayside thermosphere, where they are produced by photodissociation of  $\text{N}_2$ ,  $\text{O}_2$  and  $\text{CO}_2$ .

All these processes show a strong coupling between different atmospheric layers and between different process. As we will see, GCMs naturally include these coupling, making them valuable tools to the study of the atmosphere as a global system.

### 1.3 General Circulation Models for Mars

General circulation models have their origin in the models for meteorological prediction. They simulate the atmosphere by a 3-D grid (longitude-latitude-altitude) in which they solve by numerical methods the governing equations. They can be schematically decomposed in a core that solves the equations of the dynamics and a set of physical processes in form of approximations or parameterizations, to avoid an excessive CPU time consumption. These models include in a natural way the couplings between layers and between physical processes.

However, although these GCMs are very powerful tools, they suffer from a series of limitations that need to be considered when analyzing their results. It is not the lesser the natural unpredictability of the atmosphere, given its chaotic and almost turbulent behavior due to the non-linear character of the equations of the atmospheric physics [17], that originates a day-to-day variability difficult to predict. In many cases, temporal averages are necessary to lessen this problem. Other problems that can be mentioned are the limitations to consider processes with a scale lower than the grid size, and the excessive CPU time consumption. And, in the particular case of the Martian upper atmosphere, the lack of data to contrast the predictions of the models imposes an additional difficulty.

Several GCMs for Mars have been developed since the pioneering work of Leovy and Mintz [18]. Most of them are devoted to the study of the lower atmosphere. For example, we can cite the NASA/AMES-MGCM, used to study the temperature profiles measured by MGS [19], the GFDL Mars-GCM, that has been employed to study the thermal tides [20], or that developed by a French and British consortium, the LMD/AOPP MGCM, used for example to study the polar warming in the lower atmosphere [21]. The only thermospheric GCM until recently, the Mars Thermospheric GCM, developed at NCAR and maintained by the University of Michigan, has been used in studies of comparative terrestrial planet thermospheres [22, 23]. This model has been recently coupled to the NASA/AMES MGCM, with key fields being passed upwards from the NASA/AMES MGCM to the MTGCM, but not the other way around.

## 2 The LMD-Mars General Circulation Model

The LMD-MGCM has its origin in the model for the study of the terrestrial climate developed at the Laboratoire de Météorologie Dynamique (Paris University). To adapt this model to Mars, a new radiative transfer code [24] and a CO<sub>2</sub> condensation/sublimation scheme [25] were developed. It includes the radiative effects of CO<sub>2</sub> and dust, a number of subgrid-scale processes, processes of interchange surface-atmosphere and the seasonal cycle of CO<sub>2</sub> and H<sub>2</sub>O [21]. Originally, it extended from the surface up to about 80 km.

In the frame of a joint project of the LMD, the University of Oxford and the Instituto de Astrofísica de Andalucía (IAA, CSIC, Spain) and sponsored by the ESA, the vertical range of the model has been extended up to the upper thermosphere, becoming in this way the first Martian GCM able to study in a self-consistent way the whole atmospheric range from the surface up to the upper thermosphere. This extension has been done in two steps. First, the model was extended up to about 120 km by including the Non-Local Thermodynamic Equilibrium (NLTE) correction to the CO<sub>2</sub> NIR solar heating rate and to the cooling due to 15 $\mu$ m emissions by CO<sub>2</sub> [12]. And in a second step, it was extended up to the thermosphere by adding parameterizations for the physical processes important at these altitudes: Molecular diffusion, thermal conduction [26], photochemistry of the C, H and O families and UV heating [27]. For both extensions, a 1-D model developed at the IAA has been used to implement detailed schemes and to develop and test parameterizations to be included in the GCM.

## 2.1 The Mars Climate Database

One of the most important applications of this model is the creation of the Mars Climate Database, a compilation of statistics of the results of the LMD-MGCM [28]. This database takes into account both the diurnal and the seasonal variations. Several “scenarios” (combination of different options for the dust load and for the UV solar activity) are included to bracket the very variable conditions of the Martian atmosphere. Statistical tools to estimate the day-to-day variability are also included.

This database is currently being used by most of the active groups in the study of the Martian atmosphere, both as a reference for scientific studies and as a tool in the engineering planning of future missions. It is freely available for the community in DVD format and a simplified version can be found on-line at [www-mars.lmd.jussieu.fr](http://www-mars.lmd.jussieu.fr)

## 3 Results

### 3.1 Validation

After extending the LMD-MGCM up to the thermosphere, the first efforts were directed towards validating the model. Different strategies have been used before directly comparing with some of the scarce observational data.

First, a series of sensitivity tests were performed to check if the model reacted as physically expected to modifications in some input parameters or in the absence of some physical processes. These tests allow also for a deeper understanding of the atmosphere as an integrated system. For example, when not including the concentration changes produced by the photochemistry, an increase of temperature was found in the upper atmosphere [27]. The reason is that one of the most important effects of

the thermospheric photochemistry is the photodissociation of CO<sub>2</sub> in CO and atomic oxygen. So, when no photochemistry is included, there is more CO<sub>2</sub> than in the “nominal” simulation. Given that CO<sub>2</sub> is more efficient than CO and O in absorbing UV radiation and heating the atmosphere, more CO<sub>2</sub> implies more heating and thus a higher temperature. Another example of these tests can be found in [26], where simulations with and without parameterized orographic gravity waves are presented, concluding that these waves can serve as a coupling mechanism between the lower and the upper atmosphere, modifying the zonal mean winds and interacting with the tides.

And second, a detailed intercomparison campaign with the reference GCM of the Martian thermosphere, the MTGCM, has been performed [29]. In this intercomparison, both models were run using similar forcings and the same input conditions. Three different “scenarios” or sets of input conditions were used, designed to study the atmospheric variability with seasons and with different dust loads. A good overall agreement is found, although some local/regional differences have been identified, expected when comparing models of such a complexity. The detailed results of this intercomparison campaign will be published elsewhere.

### 3.2 Thermal and wind structure of the Martian upper atmosphere

After these validation exercises, we have exercised the model to study the thermal and wind structure of the Martian upper atmosphere. The longitudinal and latitudinal variation of the temperatures and winds predicted by the model for perihelium conditions (Southern summer) is shown in figure 2. We can clearly appreciate the shape of the terminator (the day-night separation line) and how the summer polar region is constantly illuminated. Maximum temperatures of about 400 K are found in the Equator close to the evening terminator, while the minimum temperatures ( $\approx 200$  K) are found in the equatorial region close to midnight. The winds diverge from the summer night hemisphere and converge in the Equator before midnight. The energy transported by the winds modifies the distribution of temperatures that would be expected by radiative equilibrium, as described by [23]. A thermospheric polar warming is clearly visible during the night, as observed by Mars Odyssey [14].

The balance between the different heating/cooling terms predicted by the LMD-MGCM can be found in figure 3. The UV heating is the main heating source of the Martian upper atmosphere, and it is mainly compensated at the altitude of its peak by thermal conduction, although there is an important contribution by 15  $\mu\text{m}$  cooling in lower layers, in good agreement with the predictions of the MTGCM [23].

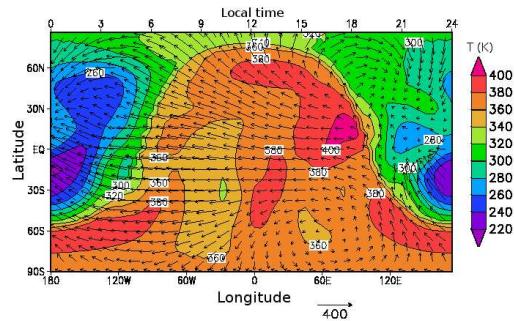


Figure 2: Temperatures (color contours) and winds (arrows) in a constant pressure layer in the upper thermosphere ( $P \approx 10^{-9}$  Pa) derived from the LMD-MGCM for perihelium conditions.

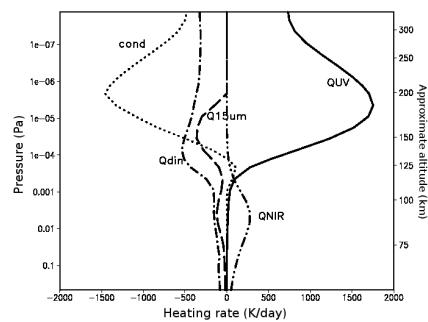


Figure 3: Balance between different heating/cooling terms for perihelium conditions.

### 3.3 Comparisons with data

In figure 4, we compare the variation of the upper thermospheric temperatures with seasons and with the solar cycle predicted by the LMD-MGCM (solid lines) with the few existing data (symbols) and with the results from the MTGCM (dashed lines), taken from [23]. Temperature is minimum for aphelium and maximum for the perihelium season, as expected. The LMD-MGCM predicts a more intense seasonal variability than the MTGCM, although it has to be taken into account that the results from the MTGCM are for a dust-free lower atmosphere. In spite of predicting reasonably well the temperatures for solar minimum conditions, the LMD-MGCM tends to overestimate the exospheric temperatures for solar medium and maximum conditions. This indicates an overestimation of the UV heating, an underestimation

of the  $15\ \mu\text{m}$  cooling or the thermal conduction, or the lack of some process. However, given the scarcity of results used for this comparison and their variability, this result needs to be further confirmed, for example with comparisons with SPICAM temperature profiles, as discussed below.

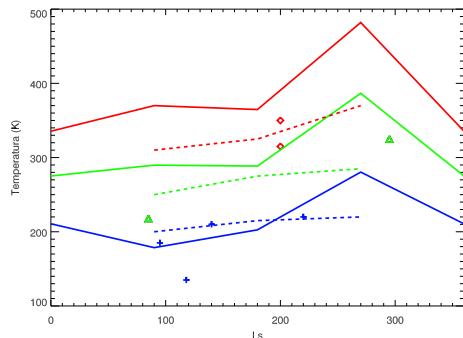


Figure 4: Variation of equatorial temperatures in the upper thermosphere at LT=15 given by the LMD-MGCM (solid lines) and by the MTGCM (dashed lines, taken from [23]) for solar minimum (blue), average (green) and maximum (red) conditions. The different symbols represent the observations: red squares for solar maximum conditions (Mariner 6 and 7), green triangles for solar average conditions (Mariner 9 and phase 2 of aerobraking of MGS) and blue crosses for solar minimum conditions (Viking Landers 1 and 2, Mariner 4 and phase 1 of aerobraking of MGS).

Maybe the most interesting application of these models is the comparison with data, that is doubly useful as a validation exercise for the model, revealing its weak points that need to be improved, and allowing a deeper understanding of the data.

A good example of this merging between data and models is the study with the LMD-MGCM of the density measured by MGS during its aerobraking. A good agreement between the data and the predictions from the model is found, although the density is underestimated by the model in the polar regions [12]. The wave structure obtained by MGS is nicely reproduced by the model, allowing to perform a Fourier decomposition of the results and confirming the importance of the non-migrating tidal components (that is, those components not directly excited by the Sun). The origin of these components is the interaction of the solar illumination with the topography and the wave-wave interactions, coupling the lower and the upper atmosphere [12].

We are currently using the model for the analysis of some other sets of data about the Martian upper atmosphere, that we briefly mention below:

SPICAM on board Mars Express has performed the first remote sensing observations of the upper Martian atmosphere, using the technique of stellar occultation [30].

About 600 profiles have been obtained during more than one Martian year, with a good latitudinal and longitudinal coverage. We are right now comparing the SPICAM profiles with results from the MCD, which will tell us for the first time the accuracy of the model at this altitude range, previously unexplored.

As mentioned above, SPICAM has also observed for the first time the NO night-glow in Mars [16]. In order to study this phenomenon with the LMD-MGCM we are currently working in the extension of the photochemical module to include the chemistry of the Nitrogen family. This will allow to simulate the production of N atoms in the dayside thermosphere and its recombination with O in the nightside mesosphere. By comparing this recombination rate with the one inferred from SPICAM NO night-glow observations, we will hopefully be able to constrain the dynamics, in particular the day-night transport.

In the future, we plan to use other recent observations to further validate our model. In particular, the emissions by CO<sub>2</sub> and O<sub>2</sub> measured by OMEGA on board Mars Express will be very valuable to constrain the concentrations predicted by the LMD-MGCM.

## 4 Summary and conclusions

General Circulation Models are very valuable tools for the study of planetary atmospheres. Given the strong coupling between the Martian lower and upper atmosphere, it is very important to study this complex system with a ground-to-exosphere model, able to study in a self-consistent way the coupling between atmospheric layers and between physical processes. With this in mind, the GCM developed at the LMD has been extended up to the thermosphere, in collaboration with the Instituto de Astrofísica de Andalucía, by adding the physical processes relevant for these altitudes.

This model has been (and is still being) carefully validated. First, a series of sensitivity tests has been performed to assess the behavior of the model when some input parameters are modified or in the absence of certain processes. Second, a detailed intercomparison with the reference GCM of the Martian thermosphere, the MTGCM, has shown a general agreement between the models. And third, we have compared the results from the model with the most recent observations of the Martian upper atmosphere.

The model reproduces the density measured during the aerobraking of MGS, although an underestimation is obtained in the polar regions. This has allowed us to confirm the importance of the non-migrating components to produce the observed wave structure. Comparisons with the exospheric temperatures measured by different spacecrafts and its variation with season and solar cycle shows that the model overestimates the exospheric temperature, at least for solar average and maximum conditions. This is a preliminary result that needs to be confirmed by comparing

with the more complete data set of temperatures in the thermosphere, obtained by SPICAM.

## References

- [1] Kliore, A., Cain, D.L., Levy, G.S. et al. 1965, *Science* 149, 1243
- [2] Kieffer, H.H., Martin, T.Z., Peterfreund, A. et al. 1977, *J. Geophys. Res.* 82, 4249
- [3] Zurek, R.W. 1992, in “Mars”, University of Arizona Press.
- [4] James, P.B., Kieffer, H.H., Paige, D.A. 1992, in “Mars”, University of Arizona Press
- [5] Hess, S.L., Ryan, J.A., Tillman, J.E. et al 1980, *Geophys. Res. Lett.* 7, 197
- [6] Anguita, F. 1998, in “Historia de Marte: Mito, exploración, futuro”, Ed. Planeta
- [7] Smith, M.D. 2006, “Proceedings of the 2nd Mars Atmosphere Modelling and Observations workshop”, ESA-CNES
- [8] Houghton, J.T. 1977, “The Physics of Atmospheres”, Cambridge University Press
- [9] Forbes, J.M., Bridger, A.F.C., Bougher, S.W. et al. 2002, *J. Geophys. Res.* 107, E115113
- [10] Bougher, S.W., Bell, J.M., Murphy, J.R. et al. 2006, *Geophys. Res. Lett.* 33, L02203
- [11] Keating, G., Bougher, S.W., Zurek, R.W. et al. 1998, *Science* 279, 1672
- [12] Angelats, M., Forget, F., López-Valverde, M.A. et al. 2004, *J. Geophys. Res.* 109, E01011
- [13] Forbes, J.M., Hagan, M.E. 2000, *Geophys. Res. Lett.* 27, 3563
- [14] Keating, G., Theriot, M., Tolson, R. et al. 2003, “Proceedings of the 1st Mars Atmosphere Modelling and Observations workshop”, CNES-ESA
- [15] Bell, J.M., Bougher, S.W., Murphy, J.R. 2007, *J. Geophys. Res.* 112, E12002
- [16] Bertaux, J.-L., Leblanc, F., Perrier, S. et al. 2005, *Science* 307, 566
- [17] James, I.N. 1994, in “Introduction to circulating atmospheres”, Cambridge University Press
- [18] Leovy, C.B., Mintz, Y. 1969, *J. Atmos. Sci.* 26, 1167
- [19] Joshi, M., Haberle, R.M., Hollingsworth, J. 2000, *J. Geophys. Res.* 105, 17601
- [20] Hinson, D.P., Wilson, R.J. 2004, *J. Geophys. Res.* 109, E01002
- [21] Forget, F., Hourdin, F., Talagrand, O. 1999, *J. Geophys. Res.* 104, 24155
- [22] Bougher, S.W., Engel, S., Roble, R.G. et al. 1999, *J. Geophys. Res.* 104, 16591
- [23] Bougher, S.W., Engel, S., Roble, R.G. et al. 2000, *J. Geophys. Res.* 105, 17699
- [24] Hourdin, F. 1992, *J. Geophys. Res.* 97, 18319
- [25] Hourdin, F., Le Van, P., Forget, F. et al. 1993, *J. Atmos. Sci.* 50, 3625
- [26] Angelats i Coll, M., Forget, F., López-Valverde, M.A. et al. 2005, *Geophys. Res. Lett.* 32, L04201

- [27] González-Galindo, F., López-Valverde, M.A., Angelats i Coll, M. et al. 2005, J. Geophys. Res. 110, E09008
- [28] Lewis, S.R., Collins, M., Read, P.L. et al. 1999, J. Geophys. Res. 104, 24177
- [29] González-Galindo, F., Bougher, S.W., López-Valverde, M.A. et al. 2006, “Proceedings of the 2nd Mars Atmosphere Modelling and Observations workshop”, CNES-ESA
- [30] Bertaux, J.-L., Korablev, O., Perrier, S. et al. 2006, J. Geophys. Res. 111, E10S90



# WHY ARE INTENSE GEOMAGNETIC STORMS SO IMPORTANT FOR HUMAN LIFE?

ELENA SAIZ, YOLANDA CERRATO, CONSUELO CID, JESÚS AGUADO  
*Departamento de Física. Universidad de Alcalá, E-28871 Alcalá de Henares, SPAIN*

**Abstract:** Increasing knowledge concerning the space environment surrounding the earth has become one of the main focuses of research. This is mainly due to the fact that the adverse conditions in near-earth space cause significant damage to technological systems and, consequently, considerable economic losses. Many types of space weather-related anomalies and failings have been identified in recent years, thus converting adverse space weather into one of the threats facing modern human technology. Therefore important efforts should be made to find technical and operational solutions to space weather problems. In this framework, the need to implement reliable real-time warning tools is evident. Meanwhile, the fewer parameters involved in making predictions, the more valuable the tools will be. The present work develops a warning procedure based on the use of the z component of the interplanetary magnetic field only. The aim of this tool is to warn of the occurrence of intense geomagnetic variations, as measured by the geomagnetic Dst index. A comparison of our results with those criteria available in the relevant literature for the occurrence of intense geomagnetic activity shows a significant improvement in alerting capability.

**Keywords:** Geomagnetic storms – Dst index – Space Weather – Hazards.

## 1 Introduction

The sun is the star which human life depends on. In quiet conditions solar wind (plasma and magnetic fields travelling together) is constantly blowing off the sun. Solar wind at the earth's orbit has a mean density of about  $4 \text{ cm}^{-3}$ , a mean velocity of about 400 km/s and a mean interplanetary magnetic field (IMF) of about 5 nT.

However, the sun is an extremely active star and disturbs the earth in several ways. Large solar flares, solar coronal mass ejections (CMEs), or high-velocity solar wind streams occur in the sun from time to time. When this happens, a considerable amount of radiation, highly energetic particle fluxes and magnetic flux is released into the interplanetary medium. Figure 1 offers a schematic representation of the linked sun-earth system.

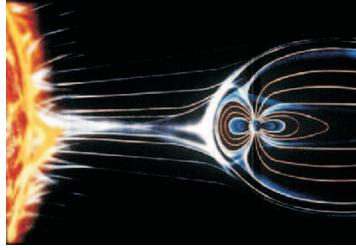


Figure 1: Picture taken from the NASA website showing the linked sun-earth system.

Interplanetary counterparts (ICMEs) of CMEs propagate outward from the sun, often with velocities of up to about 1000 km/s. As they travel faster than normal solar wind, they usually have a shock ahead, a sheath behind the shock with high density and strong magnetic field, and the ICME proper to coherent magnetic field structure.

High-velocity solar wind streams are originated in the solar coronal holes (CHs). As these streams emanating from CHs run into a slower solar wind, co-rotating interaction regions (CIRs) arise.

When these solar events, with different structures, impinge on the earth, major disturbances are detected in the earth's magnetosphere, upper atmosphere and even on the terrestrial surface; in these cases, the solar events are called geoeffective. While the shock, sheath and ICME in CME events are all effective drivers of geomagnetic activity [1], CIR events are also effective drivers but only related to medium-level activity [2].

Solar activity varies in line with the well-known 11-year cycle, a cycle which coincides with magnetic records. ICMEs are more frequent during maximum solar phase and the early part of declining phase of the cycle [3]. However, during the declining phase, where the CHs extend to low latitudes, even reaching the solar equator, CIRs are also significant sources of geomagnetic activity [4].

On the other hand, there are multiple spatial and temporal scales involved in space environment changes. Spatial scales extend from interplanetary medium, outer and inner magnetosphere, ionosphere at high and mid-latitudes up to the thermosphere. Temporal scales go from 1 to 5 days for propagation of solar wind from the sun to the earth, depending on the velocity of the interplanetary structure impinging on the terrestrial magnetosphere; a few tens of minutes for solar energetic particles events; or 8 minutes for electromagnetic radiation. However that may be, the processes, changes, and effects that take place during these important disturbances in the earth's space environment are called 'geomagnetic storms'. Nowadays, their study is of crucial importance, not only because of the intrinsic scientific value they might have, but

also more practically in so far as they lead to sequences of damage to humans in space as well as to several types of technological systems. Satellites, space-based equipment for positioning and navigation, military reconnaissance, communication and ground-based systems such as electrical power supply networks are all examples of sectors affected by disturbed conditions in space.

The term commonly used to describe the state of the space environment is 'space weather'. COST 724 action, as a contribution to the space weather community, defined that term as follows: "Space weather is the physical and phenomenological state of natural space environments. The associated discipline aims, through observation, monitoring, analysis and modelling, at understanding and predicting the state of the sun, the interplanetary and planetary environments, and the solar and non-solar driven perturbations that affect them; and also at forecasting and nowcasting the possible impacts on biological and technological systems".

Therefore, space weather involves a chain of events from eruptions on the sun, through propagation of interplanetary disturbances and interaction with the earth's magnetosphere (and other planetary magnetospheres) to their impact on the earth's environment and society. Numerous effects related to space weather have been identified in recent years ([5], [6], [7], [8]). Figure 2 shows the main systems affected when space weather disturbances take place.

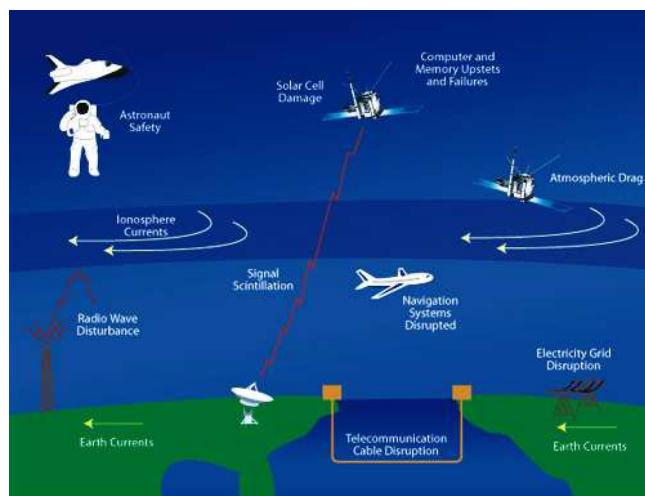


Figure 2: Main systems affected when space weather disturbances take place (taken from the NASA website).

In what follows, a brief overview is offered of the basic properties of the structure

and dynamics of the magnetosphere and of the relationships between the drivers and their consequences for both the plasma environment and technological systems.

## 2 Signatures and effects of geomagnetic storms

The terrestrial magnetosphere is a highly dynamic system controlled by non-linear interactions involving a continuous exchange of energy, mass and momentum between solar wind and magnetosphere, thus driving geomagnetic activity. In 'quiet time' conditions, the solar wind affects the magnetic structure created by the earth itself, distorting the dipolar magnetic field lines. As a result, the magnetosphere is strongly compressed on the dayside (facing the sun) and elongated or tail-like on the nightside. Consequently, the size and shape of the magnetosphere as a whole is controlled by solar wind dynamic pressure. This continuous solar wind-magnetosphere interaction leads to a complicated system of plasma convection flows and currents systems in the magnetosphere, which in turn give rise to its final topology (see Figure 3 for different parts of the magnetosphere).

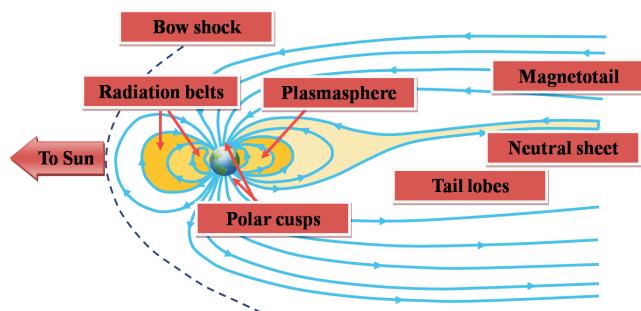


Figure 3: The different parts of the magnetosphere.

When the quiet IMF and solar wind plasma are disturbed by a solar event and impinge on the earth, the bow-shock on the dayside comes closer to the earth as a result of the dynamic pressure enhancement of the solar wind plasma (enhancement in density and velocity). During intense geomagnetic storms, spacecraft in geostationary orbits ( $6.6 R_E$ ) could be placed outside the magnetosphere in this period of time and suffer serious damage as a result.

Geomagnetic activity is the result of energy entering the magnetosphere, magnetic reconnection being the most important mechanism involved [9]. This process takes place when the IMF and terrestrial magnetic field run in opposite directions. Major geomagnetic activity is recorded when magnetic reconnection takes place at

the earth's dayside magnetopause; that is to say, the process is primarily dependent on the southward component of the IMF ( $B_z$ ) [10]. This scenario may also be described by the dawn-dusk component ( $E_y$ ) of the interplanetary electric field (IEF) by considering the solar wind as a plasma moving in a magnetic field, and therefore with a motional electric field ( $\vec{E} = -\vec{v} \times \vec{B}$ ).

As a result of reconnection at dayside magnetopause, amount of magnetic flux is carried to the magnetotail where reconnection also takes place with tail-lobes magnetic field lines. As a consequence, the current sheet gets narrower, instability may arise and plasma be released in an explosion and propagated away from the reconnection point.

The plasma heading towards the earth's inner magnetosphere augments the population of the radiation belts and penetrates into high-latitude ionosphere along field-aligned currents (FACs). Because of the gradient and curvature of the dipolar magnetic field, particles trapped in the radiation belts undergo a slow azimuthal drift, constituting the ring current. The increase in the density of drifting particles (of both interplanetary medium and ionospheric origin) during disturbed conditions produces an increase in the ring current ([11], [12]).

In the ionosphere, the present current systems undergo important changes. At mid- and low-latitudes Sq and equatorial electrojet currents exist, mainly generated by heating from solar radiation. At high-latitudes, auroral electrojets are fed by FACs, which connect the high-latitudes ionosphere and the magnetosphere [13]. During geomagnetic storms electrojets are enhanced) by convection flows from the night-side magnetosphere and energetic particles precipitating into the upper atmosphere through the polar regions. As a consequence, there is a higher degree of ionization, which gives rise to an increase in ionospheric currents and temperature, which in turn causes radial outward expansion. The atmospheric drag experienced by Lower Earth Orbiting (LEO) satellites increases, reducing their velocity and causing them to lose altitude, modify their orbit and eventually enter the atmosphere prematurely.

Global positioning systems (GPS) (made up of a satellite fleet orbiting in different planes that cross the outer radiation belt) and terrestrial radio navigation systems may be seriously affected by unexpected changes in charged particle density in the storm-time ionosphere. These changes modify the amplitude and phase of waves, generating distortion and signal intensity fluctuations, as well as gradual power losses, which can lead to loss of communication with the satellite during extreme events.

At high latitudes, solar wind particles can penetrate directly into the upper atmosphere through the polar cusps. Whether by inducing chemical reactions or direct collisions with the atmospheric constituents, precipitating keV-electrons produce excitation and subsequently radiate energy over a broad range of wavelengths (infrared, visible, ultraviolet). This light, called "aurora", can be seen at lower latitudes during geomagnetic storms. Although a pleasing aspect of magnetic storms, these energetic electrons which come into the atmosphere are harmful to spacecraft or aircraft in

polar orbits.

Geomagnetically induced currents (GICs) arise on the terrestrial surface as a consequence of magnetic field variations associated with magnetospheric and ionospheric currents. GICs flow in any conducting structure such as electrical power supply network transformers, gas and oil pipelines, undersea communications cables, telephone networks or railways. Moreover, magnetic field variations on their own affect geomagnetic studies for geological surveying, directional drillings, etc.

Solar energetic particle (SEPs) events are a hazard for satellites and space-based instrumentation [14]. If they become trapped in the inner magnetosphere's dipolar field, they populate the radiation belts and reside in the magnetosphere for extended periods [15]. When solar protons with energies in the range of several MeV reach a spacecraft, can penetrate the shielding and bury themselves within dielectric materials. When sufficient charge builds up there can be a powerful internal disruptive discharge, which can affect memory devices and sensitive electronics, causing software and tracking problems. Also, solar panel degradation and power loss may ensue. These solar protons also affect the chemistry of middle and upper atmospheres while colliding with atmospheric constituents: the precipitation of particles lead to an increase of  $\text{NO}_x$  compounds in the mesosphere and thermosphere, which in turn is a catalyst for ozone destruction. This way solar activity also affects the long-term balance of atmospheric chemistry.

Space weather also has to do with issues of human health. While the terrestrial magnetosphere and atmosphere afford adequate protection for humans living on the surface of the planet, in outer space, astronauts are exposed to dangerous doses of radiation. The penetration of high-energy particles in tissue cells causes genetic mutations, associated with increased long-term risks of inheritable genetic effects and cancer.

### 3 Monitoring the earth's space environment

Extreme conditions in solar activity, solar wind and geomagnetic disturbances can be observed by a large number of satellites and ground-based sensors. The aim is to know, from all the information available over the longest possible interval of time, what activity levels system operators may expect during geomagnetic storms occurrence and how much warning they need to be given of specific events in order to assess the impact on their systems. In a highly technology-dependent society solar activity and its consequences turn out to play a crucial role.

Any understanding of the intricately interrelated sun-earth system requires monitoring links in the different parts of the chain. In other words, there must be observation of the solar atmosphere, the interplanetary medium, the magnetosphere and the earth's surface if the task of forecasting space weather is to be accomplished.

Solar activity has been routinely recorded by ESA's SOHO mission since its launch in 1995. CME occurrence and their outward direction from the sun is recorded by solar coronagraphs such as LASCO, on board SOHO. Space weather warnings are given for those events that propagate in a direction likely to reach earth.

NASA's twin STEREO spacecraft were launched in October 2006. Their on-board instruments (an extreme ultraviolet imager, two white-light coronagraphs and a heliospheric imager) will provide data enabling analysis of the 3-D evolution of CME from birth on the sun's surface, as well as much improved geometry for ICME. That will allow longer-term space weather predictions.

However, the space weather predictions based only on solar records are not enough because the effects in the near-earth environment are strongly dependent on the orientation of IMF and the velocity of solar wind plasma when encountering the earth. From solar observations alone it is not possible deduce the polarity of magnetic field structure and plasma velocity (unlike plasma velocity near the solar surface) when nearing earth.

Details of storm intensity can be predicted when an interplanetary structure (ICME, shock, CIR, etc.) arrives at the first Lagrangian point distance (L1). There, satellites such as the ACE satellite monitor *in situ* the strength and polarity of the interplanetary magnetic field as well as the density, velocity, and temperature of solar wind plasma. The L1-earth distance makes the prediction from *in situ* data only available about 1 hour in advance.

Another scientific mission whose objective is to monitor space weather from a unique perspective is the Ulysses spacecraft. Its orbit is highly inclined to the ecliptic plane, passing then over the sun's poles. It provides measurements of IMF, plasma, solar energy particles and cosmic rays. Together with the ACE, it enables the evolution of an interplanetary structure to be studied [16].

On the other hand, as solar energetic particle events in the sun impinge rapidly on the earth, the detection of active events with solar X-ray monitors, which routinely monitor the terrestrial environment, provide nowcasts for the space environment.

The terrestrial magnetosphere has been explored by numerous spacecraft orbiting the earth, but direct measurements of the magnetospheric processes during space weather disturbances are limited to a few points in space because the region to be explored is so vast that more comprehensive coverage would require a much higher number of satellites.

Spatial missions like Cluster, launched in 2000 by ESA, are continuously monitoring different zones of the magnetosphere. Cluster is a constellation of four spacecraft in tetrahedral formation with identical instruments on board. The distance separating spacecraft can be changed, so from four point measurements three dimensional vector quantities are revealed (see, e.g., [17]).

From these observational data, precise measurements of the size and speed of the bow-shock, 3D magnetic field topology at the magnetopause, typical spatial scales of

high-speed plasma flows propagating from magnetotail to the inner magnetosphere, and the thickness of plasma sheet (obtained using the curlometer technique) have all been obtained for the first time.

Populations such as charged ring current particles and neutral atoms from the plasmasphere are highly sensitive to processes occurring during space weather events such as charge-exchange processes. NASA's IMAGE mission, with its on-board high-energy neutral atom (HENA) imager, monitors the plasmaspheric and ring current dynamics in the inner magnetosphere. This is a clear demonstration that neutral atom imaging is becoming a new tool to monitor the state of the inner regions of the magnetosphere [18].

The geomagnetic activity on the earth's surface is recorded at a variety of ground magnetometer stations. Geomagnetic indices, based on these records, have been drawn up to characterize the variability of the earth's magnetic field, for all its complexity, in a single number.

The AE index measures the horizontal component of the magnetic field at terrestrial surface at auroral latitudes. The signature of geomagnetic activity in the auroral zone, as measured by this index, is an increase of the horizontal component of the magnetic field, reflecting the enhancement of the strength of the auroral electrojet. On the other hand, the Dst index measures the horizontal component of the magnetic field at terrestrial surface at low and middle latitudes. The signature of a geomagnetic storm, as recorded by Dst index, is a large depression known as main phase, followed by a recovery phase when the magnetosphere returns to its quiet state. The geomagnetic storm intensity is classified as moderate, intense or super-intense according to the minimum value of Dst reached during the storm [10]. While the main phase is the consequence of the energy input from the solar wind to the magnetosphere, the recovery phase is the result of the physical loss processes taking place in the ring current associated with neutral particles ([19], [20], [21]).

## 4 Current state of forecasting intense geomagnetic activity

A useful tool which helps to understand the space weather processes and forecast their potential effects is numerical simulation. From magnetohydrodynamic (MHD) models (using single-fluid description to characterise solar wind-magnetosphere interaction and coupling to the ionosphere), large-scale global MHD simulations have been developed ([22], [23]) to describe magnetospheric evolution. These models provide a large-scale framework for local observations and allow global quantities to be inferred which cannot be obtained directly from observation. However, they are still far from being an accurate forecasting tool.

The first empirical attempt at forecasting was that of Gonzalez and Tsurutani,

which searched for the interplanetary cause of intense geomagnetic storms. They found that  $E_y$  (calculated as  $VB_z$ ) greater than 5 mV/m over periods exceeding 3 hours were related to intense storms. It should be noted that this forecasting procedure involves  $B_z$  and solar wind velocity. As plasma instruments can be seriously affected by enhanced solar X-ray and energetic particle fluxes, they fail more often than magnetometers; moreover, sometimes the solar wind speed exceeds the upper instrumental limits of plasma detectors. Yet during these events, there is an acute need for a reliable Dst forecast since such disturbances may be accompanied by very large geomagnetic storms. This scenario led Tsurutani and Gonzalez to look for another procedure for forecasting which was not reliant on solar wind velocity. They found that the equivalent magnetic field condition for intense geomagnetic storm was  $B_z < -10$  nT, lasting at least 3 hours.

In this section we turn to provide an overview of the ability to forecast intense geomagnetic activity using the previous criteria, that is, to provide hazards of events when the Dst index reaches a value lower than -100 nT, which corresponds to the threshold value required for a storm event to be considered intense [10]. For that purpose, Dst index data provided by the World Data Center for Geomagnetism at Kyoto have been used, while historical IMF and solar wind velocity data from ACE spacecraft have been handled as if they were real-time data. The period covered extends from the spacecraft launch, at the end of 1997, to 2006.

In the first stage, Dst data were analyzed in order to look for those times when the index fell below -100 nT. In an intense geomagnetic storm event, the Dst index is used to show values lower than -100 nT for a few days. Thus, from a forecasting point of view, all those times belong to the same event, of which warning should have been given before Dst fell below the threshold for the first time. A collection of those “first times” is the set of events that a warning procedure should forecast.

The results obtained after using the criterion of Gonzalez and Tsurutani [24] are shown in Table 1, where not only the percentage of hits or misses out of the total number of events are included, but also out of the total number of events for which data was available. Although a data gap will provide a miss hazard, from the scientific point of view only a statistical analysis of available data is acceptable.

	% out of total data available	% out of total data
Hits	49	45
Misses	51	48
False alarms	9	8

Table 1: Results of the statistical analysis of hazards of intense geomagnetic storm events following the criteria proposed by Gonzalez and Tsurutani [24]:  $E_y > 5$  mV/m for  $\Delta t \geq 3$  h

Table 2 summarises the results obtained after using the criterion of Tsurutani and Gonzalez [25]. In this case, there is no data gap, as this last criterion only involves the z-component of IMF. In both tables, those events for which the forecasting criteria were fulfilled but whose Dst index was over -100 nT are included under the term “false alarm”. On the other hand, the term “miss” is used for those events whose Dst index was below -100 nT but for which the forecasting criteria are not fulfilled. Note that the results of both cases are below 50 % of hits, which indicates that a forecasting tool based on the above procedures is not trustworthy for technological purposes.

	% out of total data
Hits	60
Misses	40
False alarms	18

Table 2: Results of the statistical analysis of hazards of intense geomagnetic storm events following the criteria proposed by Tsurutani and Gonzalez [25]:  $B_z < -10$  nT for  $\Delta t \geq 3$  h

## 5 Our proposal for warning

The foregoing results encouraged us to look for other features in solar wind that might allow intense geomagnetic activity to be forecasted more reliably. In terms of Faraday’s law, it is reasonable to think that fast  $B_z$  variations could be the signature of a sudden energy release from solar wind to the ring current, and then of a sharp decrease in Dst index. Figure 4 shows an example which supports this idea.

So, using the same database as in the previous section, we checked the warning capability of Dst index variations below -50 nT in one hour; but this time we took into account a new feature, namely, significant  $B_z$  variations over a certain time interval. As this kind of storm, with its rapidly developing main phase, is the most dangerous for technological systems, it is precisely the type of storm space weather forecasting tools should be on the lookout for. Several time scales and thresholds for  $B_z$  variations were considered with a view to optimising the warning tool, that is to say, to providing as many hits as possible with the minimum number of false alerts. Different resolution data were also been considered to the same end. The results as a function of the geoeffectiveness of the event are shown in Figure 5.

From just a quick glance at Figure 5 we can deduce that this new  $B_z$  signature is more successful than the criteria considered in the previous section, particularly when five-minute resolution data are used. In this case all the geomagnetic storms

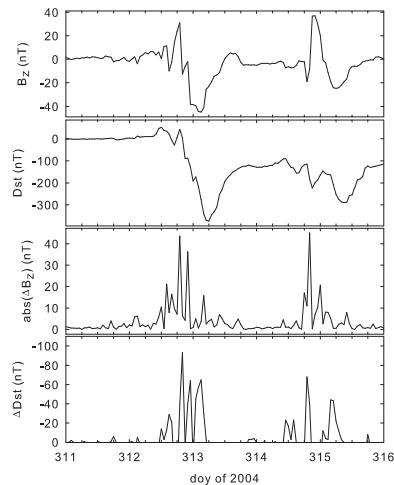


Figure 4: (From top to bottom) Interplanetary magnetic field z- component, Dst index, absolute value of  $B_z$  hourly variation and hourly Dst variation for the event happened in November, 2004.

with Dst variations lower than -75 nT in one hour were warned of. Also noteworthy is the reduction in the number of false alarms to only 4 events. As to the number of gaps, it should be noted that the reduction is due to the fact that the warning tool took no account of solar wind velocity data. Moreover, most “misses” showed fast  $B_z$  variations between 30 and 40 nT and corresponded to events in which the Dst index varied between -50 and -60 nT. Therefore, it is reasonable to suppose that these misses could be included within the accuracy range of the warning tool. However that might be, our proposal for warning yielded very good results, especially with five-minute resolution data. For this data set, the proportion of “hits” is over 70 %, which represents a major advance in space weather forecasting tools.

*Acknowledgements:* We would like to thank the World Data Center for Geomagnetism in Kyoto for providing the Dst index. We also thank the ACE MAG and SWE instruments teams and the ACE Science Center for the magnetic field and solar wind plasma data. This work has been supported by grants from the Comisión Interministerial de Ciencia y Tecnología (CICYT) of Spain (ESP 2005-07290-C02-01 and ESP 2006-08459).

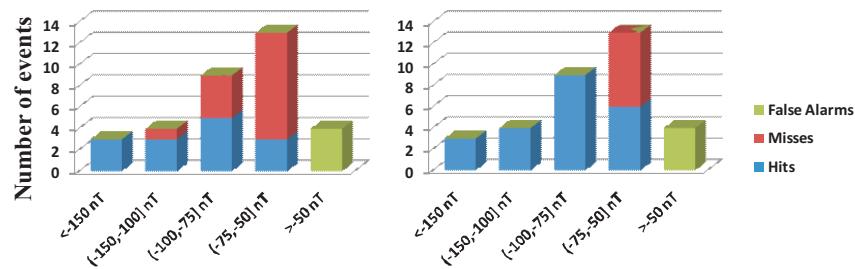


Figure 5: Results of the warning tool as a function of hourly Dst variation using  $B_z$  variations at a time interval of 3 hours as a tracer. Left panel corresponds to hourly resolution data, with threshold for  $B_z$  variations at 30 nT; right panel corresponds to five-minute resolution data with the threshold at 44 nT.

## References

- [1] Farrugia, C.J., Burlaga, L.F., Lepping, R.P. 1997, in “Magnetic Storms”, Geophys. Monogr. Ser., 98, p. 91, B.T. Tsurutani et al., eds. (AGU, Washington D.C.)
- [2] Tsurutani, B.T., Gonzalez, W.D. 1987, Planet. Space Sci. 35, 405
- [3] Bothmer, V., Schwenn, R. 1998, Ann. Geophys. 16, 1
- [4] Paulikas, G.A., Blake, G.B. 1979, in “Quantitative modeling of magnetospheric processes”, Geophys. Monogr., Vol. 21, p. 180, B.P. Olson, ed. (AGU, Washington D.C.)
- [5] Lundstedt, H. 1992, “Proceedings of Solar-Terrestrial Workshop”, NOAA, 607
- [6] Boteler, D. 1993, “IEEE PES Winter Meeting”
- [7] Viljanen, A., Pirjola, R. 1994, Sur. Geophys. 14, 308
- [8] Lanzerotti, L.J. 2001, “Space Weather effects on technologies”, AGU
- [9] Dungey, J.W. 1961, Phys. Res. Lett. 6, 47
- [10] Gonzalez, W.D., Joselyn, J.A., Kamide, Y., Kroehl, H.W., Rostoker, G., Tsurutani, B.T., Vasyliunas, V.M. 1994, J. Geophys. Res. 99, 5771
- [11] Valtonen, E. 2005, in “Space Weather Effects on Technology”, Lecture Notes Phys., 656, p. 241, K. Scherer, H. Fichtner, B. Heber et al., eds. (Springer-Verlag, Berlin Heidelberg)
- [12] Otto, A. 2005, in “The magnetosphere”, Lecture Notes Phys. 656, 133, K. Scherer, H. Fichtner, B. Heber et al., eds. (Springer-Verlag, Berlin Heidelberg)
- [13] Devendraa Siingh, Singh, R.P., Kamra, A.K. et al. 2005, J. Atmos. Sol.-Terr. Phys. 67, 637
- [14] Baker, D.N. 2000, IEEE Trans. Plasma Sci. 28, 2007

- [15] Hudson, M.K., Kress, B.T., Mazur, J.E. et al. 2004, *J. Atmos. Terr. Phys.* 66, 1389
- [16] Rodriguez, L., Zhukov, A.N., Dasso, S. et al. 2007, *Ann. Geophys.* 25, 1
- [17] Escoubet, C.P., Fehringer, M., Goldstein, M. 2001, *Ann. Geophys.* 19, 1197
- [18] Burch, J.L., Mende, S.B., Mitchell, D.G. et al. 2001, *Science* 291, 1
- [19] Jordanova, V.K., Kistler, L.M., Kozyra, J.U. et al. 1996, *J. Geophys. Res.* 101, 111
- [20] Kozyra, J.U., Fock, M.C., Sanchez, E.R. et al. 1998, *J. Geophys. Res.* 103, 6801
- [21] Kozyra, J.U., Jordanova, V.K., Home, R.B., Thorne, R.M. 1997, in “Magnetic Storms”, *Geophys. Monogr. Ser.* 98, B. T. Tsurutani et al., eds. (AGU, Washington D.C.)
- [22] Lyon, J.G., Fedder, J.A., Mobarry, C.M. 2004, *J. Atmos. Sol.-Terr. Phys.* 66, 1333
- [23] Hanhunen, P. 1996, in “Environment Modelling for Space-based Applications”, *ESA Conference Proceedings, SP-392*, 233, A. Hilgers, T.-D. Guyenne, eds. (ESA Pub. Div., Noordwijk, Netherlands)
- [24] Gonzalez, W.D., Tsurutani, B.T. 1987, *Planet. Space Sci.* 35, 1101
- [25] Tsurutani, B.T., Gonzalez, W.D. 1995, *J. Atmosph. Sol. Terr. Phys.* 57, 1369



# THE PHYSICS OF THE HII REGIONS: THE PHOTOIONIZATION EQUILIBRIUM IN THE HII REGIONS OF M51

LEONEL GUTIÉRREZ<sup>1,2</sup>, JOHN E. BECKMAN<sup>1,3</sup>

*1 Instituto de Astrofísica de Canarias. C/Vía Láctea s/n, La Laguna, 38200, Tenerife (SPAIN)*

*2 Universidad Nacional Autónoma de México, Apartado Postal 877, Ensenada 22800, Baja California (México)*

*3 Consejo Superior de Investigaciones Científicas (Spain)*

**Abstract:** A general overview of the physical properties of the HII regions is presented here, emphasizing the various general models made to explain their structure and behaviour. The Lyman continuum effective extinction, representing a measure of the extinguished or lost fraction of the ionizing flux, is explained briefly, and the results of a study performed on this matter are presented. This study, based on a sample of about 200 HII regions of the galaxy M51 using high resolution images acquired with the HST, suggests a highly inhomogeneous structure.

**Keywords:** HII regions – Star formation rate – Extinction.

## 1 Introduction

A simple measure of the star formation rate (SFR) in a galaxy is the H $\alpha$  luminosity emitted by the ionized gas in the galaxy. However, according to the evidence, the H $\alpha$  luminosity is not completely satisfactory as a star formation indicator. Several factors, notably the extinction produced by the dust both of the H $\alpha$  luminosity and of the Lyman continuum before it can ionize the gas, do not permit an exact measurement of the SFR using only the H $\alpha$  luminosity. For highly obscured star-forming regions, a better tracer of SFR could be the total infrared emission (IR) [1], mainly the mid-IR. This has been confirmed by several studies carried out using data at 7 and 15  $\mu m$  [2], at 6.7, 12, and 15  $\mu m$  [3] and 24  $\mu m$  [4, 5, 6, 7]. The results of Relaño et al. [7] show that the 24  $\mu m$  luminosity can be used as a SFR tracer in individual HII regions, but for the integrated emission from entire galaxies it is necessary to take into account the metallicity of the galaxy. In addition, Calzetti et al. [8] found that the 8  $\mu m$  emission shows strong dependence on metallicity and on the size of the region being

measured; also, they found that the  $24\ \mu\text{m}$  emission bears a nonlinear relation with the  $\text{Pa}\alpha$  emission (another useful tracer of SFR, but  $\sim 8$  times weaker than  $\text{H}\alpha$ ). On the other hand, even assuming that an important fraction of the luminous energy emitted by recently formed stars is reprocessed by dust in the infrared, there is a possible problem: the evolved stellar population also heats the dust giving an overestimate of the SFR. So, the best SFR tracer may well be a combination of  $24\ \mu\text{m}$  and  $\text{H}\alpha$  emission. However, in any case it is important to understand the efficiency of the ionization process in HII regions and know the behaviour of the dust in a galaxy in order to estimate its effects on the determination of the SFR.

M51 is a spiral galaxy (NGC 5194) interacting with its companion NGC 5195. Its essentially face-on orientation ( $i \sim 40$  degrees) allows us to observe its structure with excellent detail and with a minimum of obscuration by interstellar dust. This makes it one of the best extragalactic systems for studying star formation.

The statistical properties as well as the physical properties of the HII regions of M51 have been studied by various authors, from Carranza et al. [9], to Calzetti et al. [5]. However, the aim of the present work is, using high resolution images, to quantify the missing fraction of Lyman continuum photons that are extinguished before ionizing the gas or simply escape from the region. These results can give us evidence about whether the gas and dust distribution are homogeneous.

To do that, a set of HII regions (about 200), with distances from the galaxy centre of between 5 and 11 kpc, has been analysed identifying by their colours the star clusters which are responsible for the gas ionization there. The Lyman continuum luminosity which should be produced by the clusters was calculated, as well as the  $\text{H}\alpha$  luminosity which would be emitted if the gas were ionized in ideal conditions. Comparing this luminosity with the measured value corrected for extinction, we found values for the Lyman continuum effective extinction that reinforce the idea that the gas and dust distribution are very inhomogeneous.

## 2 The HII regions and models of their structure

When big clouds of interstellar gas collapse, they give rise to the formation of stars and star clusters. Once the stars are formed in the interior of a giant molecular cloud, many of them, the younger and more massive stars ( $M_* > 10 M_\odot$ ) with high surface temperatures ( $3 \times 10^4 \text{K} < T < 5 \times 10^4 \text{K}$ ), are stars of spectral type O and B which are generally grouped in “OB associations”. They are intense sources of UV radiation and release significant fluxes of photons with energies greater than 13.6 eV (the ionization potential of hydrogen) ionizing the surrounding hydrogen. This ionized gas becomes an HII region, usually surrounded by the neutral hydrogen shielded from the UV radiation. The higher the temperature and luminosity of the ionizing stars, the more luminous is the UV radiation and the higher is the ionization degree of the region [10,

11].

Hydrogen is the main constituent of the interstellar gas, hence it is the main component of the HII regions. However, it is possible to find in these regions other heavier elements in ionized states, mainly He, O, N, C and Ne simply or multiply ionized.

The electrons (or, better, photoelectrons) released in the ionization process, with local densities in the range  $10 - 10^4 \text{ cm}^{-3}$ , have kinetic energies given by the difference between the energy of the ionizing photon and the ionization potential of hydrogen. These electrons collide with other electrons and ions distributing their kinetic energy maintaining a maxwellian velocity distribution which defines the electron temperature of the HII region. Typically, this temperature is of order 7000 – 13000 K.

The electron-ion collisions can excite atoms of other types in the region, which emit in forbidden spectral lines. Among the more representative ions, excited in this way, are  $\text{O}^+$ ,  $\text{O}^{++}$ ,  $\text{N}^+$ ,  $\text{S}^+$ ,  $\text{S}^{++}$ , emitting in the lines [OII], [OIII], [NII], [SII] and [SIII], respectively. Moreover, the recombination of these “thermalized” electrons with the excited ions (generally  $\text{H}^+$  and  $\text{He}^+$ ), followed by downward transitions in the energy level, give rise to the emission of “recombination lines” which, with the forbidden lines, dominate the emission spectrum of an HII region.

Hydrogen is globally ionized in an HII region and accounts for more than 75% of the mass of the region. So we can consider the ionization structure of the hydrogen as a first approximation to the global structure of the region. Therefore, let us idealize the region as a sphere with uniform density and volume  $\frac{4}{3}\pi R_s^3$ , in which, in a unit time, the number of ionizing photons ( $Q_o$ ) emitted by the stars lying inside is equal to the number of electron-ion recombinations. The flux falls with the distance to the ionizing stars due to geometric dilution and absorption along the path of the photons, until it reaches a level where the ionization and recombination rates are in equilibrium. This radius  $R_s$  corresponds to this “borderline” and it is named the *Strömgren Radius* [10]. To compute this radius, we can assume that the region is optically thick and that the “on-the-spot” assumption is valid [11], i.e. there exists a high probability for the ionizing photons to be absorbed so that the recombinations to the ground state release photons immediately absorbed by other atoms nearby, with no net effects on the global ionization of the region (this is the so called “case B”). So, if  $N_p$  and  $N_e$  are the ion and electron densities, respectively, assumed constant inside the region, the total number of recombinations to excited levels will be given by  $\frac{4}{3}\pi R_s^3 N_p N_e \alpha_B$ , where  $\alpha_B$  is the recombination coefficient to all the excited levels of the hydrogen.

Because we assume that the hydrogen is fully ionized, then  $N_e = N_p \sim N_H$  and

$$Q_o = \frac{4}{3}\pi R_s^3 N_H^2 \alpha_B \quad (1)$$

Spectral type	T(K)	log( $Q_o$ ) (photons/sec)	$R_s$ (pc)	
			$N_H \sim 1cm^{-3}$	$N_H \sim 10cm^{-3}$
O5	40900	49.22	80	17
O6	38900	48.99	67	14
O7	36900	48.75	56	12
O8	34900	48.44	44	9
O9	32900	48.06	33	7
O9.5	31900	47.88	29	6

Table 1: Strömgren radius  $R_s$  as a function of the spectral type of the star and of the density of the surrounding hydrogen.

where  $N_H$  is the hydrogen density in atoms  $cm^{-3}$ . Then, the Strömgren radius is given by

$$R_s = \left( \frac{3Q_o}{4\pi N_H^2 \alpha_B} \right)^{\frac{1}{3}} \quad (2)$$

The radius  $R_s$  defines a geometrically well defined transition from a sphere of fully ionized hydrogen to the neutral hydrogen surrounding it. Several values for  $R_s$  are shown in Table 1 assuming that the ionizing source is an individual star and considering the values for  $Q_o$  given by Martins et al. [12]. The thickness of the transition zone is approximately equal to one mean free path of an ionizing photon, i.e.  $d \sim \frac{1}{a(\nu)N_H}$ , where  $a(\nu)$  is the ionization cross section for H by photons with energy  $h\nu$ , which must be higher than the 13.6 eV threshold. Typically  $a(\nu) \sim 6 \times 10^{-18} cm^2$ . So, with  $N_H \sim 1 cm^{-3}$ , the thickness is  $d \sim 0.05 pc$ , and with  $N_H \sim 10 cm^{-3}$ ,  $d \sim 0.005 pc$ , small compared with the typical sizes of the HII regions (see Table 1).

The degree of ionization of the other chemical elements in the region depends on the distance to the ionizing star and on the effective stellar temperature. The second most abundant element after hydrogen is helium. In this case, the first ionization potential is 24.6 eV and the second one is 54.4 eV, so that only a very small quantity of doubly ionized He exists in HII regions because even the hottest O stars emit few photons more energetic than 54.4 eV. Photons with energy in the range 13.6 – 24.6 eV ionize the hydrogen and those with energy greater than 24.6 eV can ionize the helium too. Depending on the ionizing star temperature, also ions of other less abundant elements such as  $O^{++}$ ,  $N^{++}$  and  $Ne^{++}$  may be found.

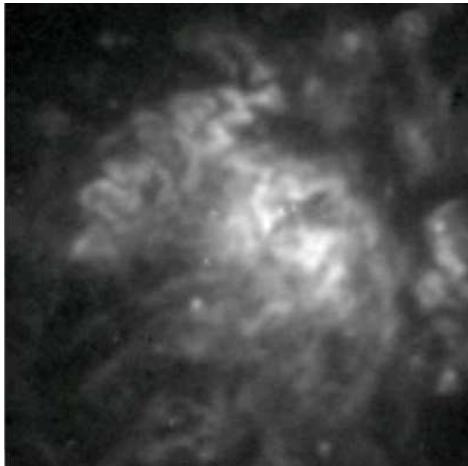


Figure 1: One of the HII regions of M51 showing its complex structure. This image spans over  $7.5'' \times 7.5''$  ( $\sim 300pc \times 300pc$ ).

## 2.1 Ionization bounded and density bounded HII regions

The volume of the ionized gas surrounding a hot star, assumed spherical and uniform, is determined by the radius  $R_s$  inside which the ionizing photons are absorbed, and therefore is proportional to its ionizing luminosity and inversely proportional to the gas density squared as we showed in the previous section (Eq. 2). The radius of these regions is limited by the rate of ionizing photons ( $Q_o$ ) emitted by the stars and not by the size of the HI cloud [10]. This cloud remains neutral outside the ionized region. These regions are known as *ionization bounded regions*. However, observations have shown that there exist some regions where there is not enough gas to absorb the ionizing radiation released by the stars. These regions are known as *density bounded regions*.

## 2.2 Filling factor

The idealization shown above works well as a practical guide to understand the phenomenon in a global way. Nevertheless, the observations show that the structures are more complex than that ideal sphere (see Figure 1). Almost all nebulae show some structure: clumps, filaments, envelopes, density gradients, etc.

Seaton and Osterbrock [13] inferred the existence of strong density fluctuations from the differences in the densities derived from ratios of forbidden lines and from the surface brightnesses measured in  $H\beta$ .

Osterbrock and Flather [14] studied in detail the density distribution in the Orion Nebula and showed that the densities derived from fluxes at radio wavelengths were considerably less than those derived from the ratio of the forbidden lines [OIII]  $\lambda 3729/\lambda 3726$ , suggesting that only a fraction of the nebula ( $\epsilon \sim 0.03$ ) is filled with high density fully ionized matter (clumps) and the rest is almost empty. This introduced the “filling factor” concept, defined by

$$\epsilon = \frac{\langle N_e \rangle_{rms}^2}{N_{e,fl}^2} \quad (3)$$

where  $N_{e,fl}$  is the electron density determined from the forbidden lines ratio and  $\langle N_e \rangle_{rms}$  is the *rms* electron density determined from one Balmer line or from the radio continuum flux.

### 2.3 Inhomogeneous models

Other models came later [15, 16, 17, 18], some of them really qualitative, but emphasizing the inhomogeneity of the gas and dust distribution in the nebulae. In a semi-quantitative model, Giammanco et al. [17], suggested that the clumps of gas may well be optically thick, so that only a fraction of the clump is ionized. The depth of the ionized portion is then inversely proportional to the square of the distance from the star, so that in all but the nearest clumps only a thin “skin” of any clump is ionized (see Figure 2). A consequence of this model is that considerable neutral gas may exist in an HII region. Another consequence is that the distinction between *ionization bounded regions* and *density bounded regions* is blurred and, generally, a fraction of the ionizing photons could escape, depending on the physical parameters of the region. In this sense every HII region may be density bounded.

Based on this model and defining the *Lyman continuum effective extinction* as  $A_{Ly,eff} = -2.5 \log(Q(r)/Q_o)$ , where  $Q(r)$  is the Lyman continuum luminosity measured at a distance  $r$  from the ionizing star, it is possible to show that if the characteristic filling factor of a region is constant, then the ratio  $A_{Ly,eff}/A_V$  must be proportional to  $1/r$ . In the same way, if the filling factor varies, increasing toward the centre as  $\epsilon = \left(\frac{\epsilon_o r_o}{r}\right)$ , for  $r$  much greater than a minimum radius  $r_o$ ,

$$\frac{A_{Ly,eff}}{A_V} \sim \frac{1}{\log\left(\frac{r}{r_o}\right)} \quad (4)$$

If the filling factor grows with the radius as  $\epsilon = \epsilon_o \left(1 - \frac{r_o}{r}\right)$ , for large values of  $r$ ,

$$\frac{A_{Ly,eff}}{A_V} \sim \frac{1}{\left(\frac{r}{r_o}\right) - \log\left(\frac{r}{r_o}\right)} \quad (5)$$

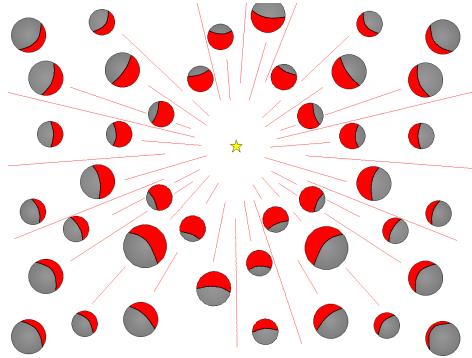


Figure 2: Illustration of the model by Giammanco et al. [17]. The gas is distributed in clumps where only a thin shell is ionized. The thickness of the shell is inversely proportional to the square of the distance to the ionizing star.

### 3 The data used in this work

M51 (NGC 5194) and its companion NGC 5195 were observed in January 2005 in 4 filters ( $B$ ,  $V$ ,  $I$  and  $H\alpha$ ) with the ACS (Advanced Camera for Surveys) on board the HST (Hubble Space Telescope). The data were processed with the automatic ACS pipeline [19] and corrected for bias, dark current and flat-field. The different defects of the detector (bad columns, hot pixels, charge transfer efficiency defects, saturated pixels, etc.) were corrected by the Multidrizzle software, a process that also corrects the filter-specific geometric distortion on each image and removes cosmic rays and other undesirable features [20].

From the images acquired with broadband filters we selected all of the sources associated with an HII region. The photometric measurement was done using aperture photometry in the three images  $B$ ,  $V$ ,  $I$ . With the measured values of luminosity, the absolute magnitudes were then determined, assuming a distance modulus of 29.62 (8.4 Mpc; [21]). Then, colours  $B-V$  and  $V-I$  were calculated to construct colour-colour diagrams and, from there, we could estimate the  $V$ -band extinction for each source. We then assigned to the region the median of the individual  $V$ -band extinction, corrected for  $H\alpha$  using the extinction curve of Cardelli et al. [22]

In the case of the  $H\alpha$  images, we subtracted the continuum emission using a linear combination of the  $V$  and  $I$  images, properly scaled as described by Knapen et al. [23]. We measured the  $H\alpha$  luminosity in  $\text{ergs s}^{-1}$ , using the conversion factor  $\text{PHOTFLAM} = 1.99918 \times 10^{-18} \text{ ergs s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ , the bandwidth of the filter-detector system  $\text{PHOTBW} = 37.15 \text{ \AA}$  and the distance to M51 as  $2.59 \times 10^{25} \text{ cm}$ .

## 4 Calculation of Lyman continuum and the predicted $H\alpha$ luminosities

Assuming that the absolute magnitude of an O7V star is  $V_{O7V} = -4.63$  and that the Lyman continuum luminosity emitted by an O7V star is  $Q_{O7V} = 5.62 \times 10^{48}$  photons  $s^{-1}$  [12], we computed the total equivalent number of O7V stars ( $N_{O7V}$ ) in every cluster [24] (i.e. the number of O7V stars that would produce the same luminosity, assuming a Salpeter [25] initial mass function), as well as the Lyman continuum luminosity produced by all of the clusters in a given HII region

$$Q_o = \sum_i N_{O7V}^i \times Q_{O7V}. \quad (6)$$

On the other hand, assuming a dust free ideal region, optically thick to the Lyman continuum radiation [11], the number of ionizing photons to be emitted per second by a stellar population inside a region is given by

$$Q'_o = 2.2 \frac{\lambda_{H\alpha}}{hc} L(H\alpha) \quad (7)$$

where  $L(H\alpha)$  is the  $H\alpha$  luminosity in  $\text{ergs s}^{-1}$  and  $\lambda_{H\alpha}$  is  $6563\text{\AA}$ . This is  $Q'_o = 7.27 \times 10^{11} L(H\alpha)$ . So, we could compare the two estimated values,  $Q_o$  and  $Q'_o$ .

## 5 Results and conclusions

In this work, the analysis is limited to HII regions localized in or around one of the arms of M51, and the minimum distance to the galactic centre is  $\sim 5$  kpc to avoid problems due to overlapping. In general, we did not find any evident correlation between the extinction calculated and the galactocentric distance in the analysed radial interval (see Figure 3). The mean  $V$  extinction value of the analysed regions is 1.39 mag, with a minimum 0.4 mag and a maximum 3.6 mag. The Galactic extinction, assumed as  $A_{V,G} = 0.116$  mag [26], has been subtracted off to give these values.

Based on the photometric data, we estimated the Lyman continuum emitted by the ionizing sources, assuming an equivalent number of O7V stars and determining the  $H\alpha$  luminosity that they should produce in ideal conditions. We compared these values with the extinction corrected  $H\alpha$  luminosity measured from the image (see Section 4). The ratio between the *missing*  $H\alpha$  luminosity and the *predicted*  $H\alpha$  luminosity is on average 0.85.

However, the value of the *Lyman continuum effective extinction*, as defined above  $A_{Ly,eff} = 2.5 \log(L_{H\alpha,Ly}/L_{H\alpha})$  [27], where  $L_{H\alpha,Ly}$  is the  $H\alpha$  luminosity predicted from the broadband clusters luminosity, and  $L_{H\alpha}$  extinction corrected measured  $H\alpha$  luminosity, is in the range 0.0 – 4.3 mag (with a mean value 1.73 mag), assuming

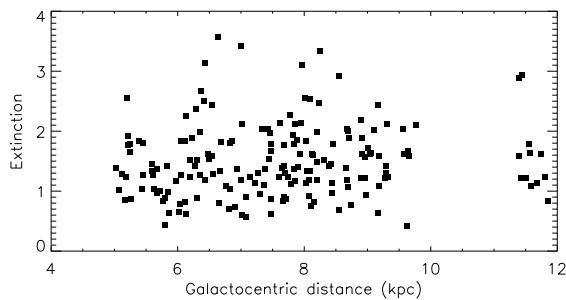


Figure 3: Mean  $V$ -band extinction in magnitudes as a function of the galactocentric distance.

that all the missing flux is due to  $A_{Ly,eff}$ . The mean value of the normalized Lyman continuum effective extinction,  $A_{Ly,eff}/E(B-V)$ , is 5.5, in the same range as the values of Hill et al. [27]. Plotting this ratio as a function of the region size it is possible to see that the values are limited by an upper envelope (see Figure 4), which shows a slow decline with size.

Some authors have suggested that up to 50 % of the Lyman continuum simply escapes [28, 29], and the lower envelope in the plot in Figure 4 suggests that there may well be a Lyman continuum photon escape factor which increases with the HII region size. Assuming an escape factor as a function increasing with the diameter, with values ranging from 0 to 75%, we applied a correction to the data and derived the plot shown in Figure 5.

This behaviour can be explained only in terms of a highly inhomogeneous gas and dust distribution. Finally, note that the shape of the plot suggests the superposition of a family of functions decreasing with increasing diameter. Could these functions have the shape of the functions presented in section 2.3 for  $\frac{A_{Ly,eff}}{A_V}$ ? If so, these results would strongly support the clumpy model from Giammanco et al. [17]

#### *Acknowledgements:*

The work presented here was carried out with the support of projects AYA2004-08251-CO2-01 and AYA2007-67625-CO2-01 of the Spanish Ministry of Education and Science, and P3/86 of the Instituto de Astrofísica de Canarias. LG thanks DGAPA-UNAM, México, for support and the Instituto de Astrofísica de Canarias for hospitality.

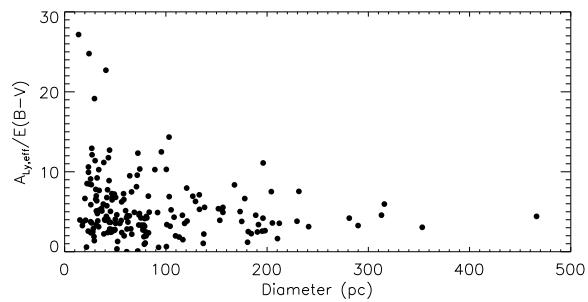


Figure 4: Normalized Lyman continuum effective extinction as a function of the diameter of the region. An upper envelope decreasing with the size and a lower envelope, slowly increasing can be observed.

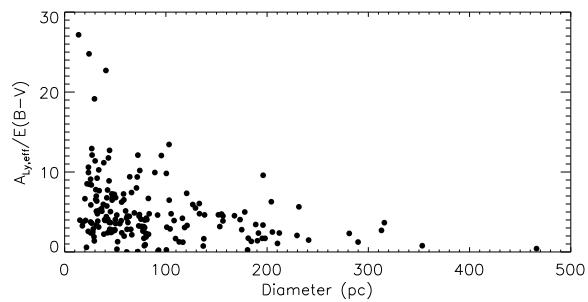


Figure 5: Normalized Lyman continuum effective extinction as a function of the diameter of the region, where the data has been corrected assuming an escape factor increasing with size to a maximum of 75%.

## References

- [1] Kennicutt, R.C. 1998, *ARA&A* 36, 189
- [2] Förster Schreiber, N.M., Roussel, H., Sauvage, M., Charmandaris, V. 2004, *A&A* 419, 501
- [3] Chary, R., Elbaz, D. 2001, *ApJ* 556, 562
- [4] Rieke, G.H., Young, E.T., Engelbracht, C.W., et al. 2004, *ApJS* 154, 25
- [5] Calzetti, D., Kennicutt Jr., R.C., Bianchi, L., et al. 2005, *ApJ* 633, 871
- [6] Pérez-González, P.G., Kennicutt, R.C., Gordon, K.D., et al. 2006, *ApJ* 648, 987
- [7] Relaño, M., Lisenfeld, U., Pérez-González, P.G. et al. 2007, *ApJL* 667, 141
- [8] Calzetti, D., Kennicutt, R.C., Engelbracht, C.W., et al. 2007, *ApJ* 666, 870
- [9] Carranza, G., Crillon, R., Monnet, G. 1969, *A&A* 1, 479
- [10] Strömgren, B. 1939, *ApJ* 89, 526
- [11] Osterbrock, D. 1989, in “Astrophysics of Gaseous Nebulae and Active Galactic Nuclei”, Mill Valley: University Science Books
- [12] Martins, F., Schaerer, D., Hillier, D.J. 2005, *A&A* 436, 1049
- [13] Seaton, M.J., Osterbrock, D.E. 1957, *ApJ* 125, 66
- [14] Osterbrock, D., Flather, E. 1959, *ApJ* 129, 260
- [15] Cantó, J., Raga, A., Steffen, W., Shapiro, P.R. 1998, *ApJ* 502, 695
- [16] O’Dell, C.R., Peimbert, M., Peimbert, A. 2003, *AJ* 125, 2590
- [17] Giammanco, C., Beckman, J., Zurita, A., Relaño, M. 2004, *A&A* 424, 877
- [18] Gahm, G.F., Grenman, T., Fredriksson, S., Kristen, H. 2007, *AJ* 133, 1795
- [19] Hack, W.J., Greenfield, P. 2000, “ASP Conf. Ser., Vol. 216, *Astronomical Data Analysis Software and Systems IX*”, eds. N. Manset, C. Veillet, and D. Crabtree (San Francisco: ASP), 433
- [20] Mutchler, M., Beckwith, S.V.W., Bond, H.E., et al. 2005, “Hubble Space Telescope multi-color ACS mosaic of M51, the Whirlpool Galaxy”, *Bulletin of the American Astronomical Society*, Vol. 37, No.2 (<http://www.aas.org/publications/baas/v37n2/aas206/339.htm>)
- [21] Feldmeier, J.J., Ciardullo, R., Jacoby, G.H. 1997, *ApJ* 479, 231
- [22] Cardelli, J.A., Clayton, G.C., Mathis, J.S. 1989, *ApJ* 345, 245
- [23] Knapen, J.H., Stedman, S., Bramich, D.M., Folkes, S.L., Bradley, T.R. 2004, *A&A* 426, 1135
- [24] Vacca, W.D. 1994, *ApJ* 421, 140
- [25] Salpeter, E.E. 1955, *ApJ* 121, 161
- [26] Schlegel, D.J., Finkbeiner, D.P., Davis, M. 1998, *ApJ* 500, 525
- [27] Hill, J.K., Waller, W.H., Cornett, R.H., et al. 1997, *ApJ* 477, 673
- [28] Oey, M.S., Kennicutt, R.C. 1997, *MNRAS* 291, 827
- [29] Relaño, M., Peimbert, M., Beckman, J. 2002, *ApJ* 564, 704



# THE FUNDAMENTAL CYCLOTRON LINE IN 4U 1538–52

JOSÉ J. RODES–ROCA , JOSÉ M. TORREJÓN , J. GUILLERMO BERNABÉU  
*Departament de Física Enginyeria de Sistemes i Teoria del Senyal, Universitat  
d'Alacant, E-03080 Alacant, SPAIN*

**Abstract:** We present pulse phase averaged spectra of the high mass X-ray binary pulsar 4U 1538–52/QV Nor. Observations of this persistent accreting pulsar were made with the Rossi X-ray Timing Explorer (RXTE) .

We study the variability of cyclotron resonant scattering feature (CRSF or simply cyclotron line) in the high energy spectra of this binary system . We show that the parameters of the CRSF are correlated. The first one is, as suggested by theory, between the width and the energy of the cyclotron line. The second one is between the relative width (defined as  $\sigma_c/E_c$ ) and the optical depth of the cyclotron line. We discuss these results with studies of other X-ray pulsars and their implications on the line variability.

**Keywords:** X-rays – Magnetic Fields – Accretion – Cyclotron Lines.

## 1 Introduction

X-ray binary systems are a perfect astrophysical laboratory to study the behaviour of matter and its interaction in extreme conditions of temperature, density, gravity and its magnetic fields. These physical properties are impossible to achieve in terrestrial laboratories. X-ray binaries are made up of a 'normal' (e. g. main sequence) companion star and a compact object. The compact object can be either a neutron star (NS) or a black hole (usually referred to as black hole candidate, BHC). Some of the observational evidence indicate that X-ray emission is generated by the accretion of material from the companion star onto the compact object. In these circumstances, matter falling onto a compact star releases gravitational potential energy, heats the matter and generates X-radiation.

In order to understand the emission properties of an accreting X-ray binary source, we need to know:

- the nature of the compact object, NS or BHC;

- the strength and geometry of the magnetic field, when the compact object is a NS;
- the geometry of the accretion flow from the companion to the compact object;
- the mass accretion rate;
- the mass of the system.

Taking the mass of the companion star into account, X-ray binaries can be classified into two main categories: High Mass X-ray Binaries (HMXB) and Low Mass X-ray Binaries (LMXB). These sources present a wide variety of phenomena, from quasi-periodic oscillations (QPOs) to X-ray outbursts. However, when binary systems contain a white dwarf (WD) and a low mass companion star, they are not called LMXB although X-ray emission is present. These sources are called Cataclysmic Variables (CVs) because they show very large variations in their brightness and are also fairly faint in X-rays.

Although there are some intermediate mass X-rays binaries (IMXB) (e.g., XTE J1819–254, GRO J1655–40, 4U 1543–475), in [1] they suggest that the majority of the LMXB systems may have descended from IMXB systems. There are some general books on X-ray binary systems as [2, 3, 4, 5] or reviews on this topic as [6, 7, 8] that can be useful to learn about X-ray astronomy. In the rest of this section we only consider HMXB systems.

In HMXB systems the companion is an O, B or Be type star with  $M \geq 10 M_{\odot}$  and the X-ray emission is produced by capture of material from the stellar wind by the compact object or through Roche-lobe overflow that can also be a supplement to the mass transfer rate in HMXB systems [2] (in Figure 1 we can see an example of both scenarios of mass transfer). Although the compact object in HMXBs may be a black hole, as in the case of Cyg X-1 (the first X-ray source discovered in the constellation Cygnus), they usually contain a NS .

Most HMXBs fall in one of the two main subgroups: those in which the primary has evolved away from the main sequence and becomes a supergiant (SG/X-ray binary) and those in which the primary has not reached the supergiant phase and is characterized by Balmer and HeI emission lines (Be/X-ray binary) .

In general, SG/X-ray binary systems are sources of persistent X-ray emission . At some stage of the binary evolution, the supergiant star ejects much of its mass in the form of a stellar wind; the NS can pick up some of these particles gravitationally and becomes a weak X-ray source. But, in the course of star evolution, another possibility may be an increase in radius of the companion star becoming a supergiant; then the outer layers of its envelope escape from the Roche lobe of the supergiant star through the internal Lagrangian point and becomes a bright X-ray source.

The most numerous class of HMXBs are Be/X-ray binaries, called also hard X-ray transients (see the catalog of [10]). The companion star is an Oe or Be star, which

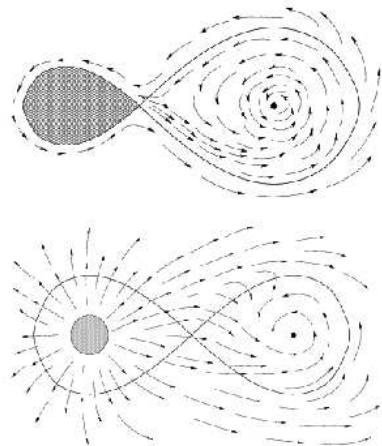


Figure 1: Mass transfer in a binary system. Top: When a massive star evolves the outer layers begin expanding and the star becomes a supergiant. If the star exceeds its Roche lobe, the material outside the Roche lobe is no longer gravitationally bound to the star. This material can now be captured by the compact object or expelled into the interstellar medium. Bottom: When the optical companion is an O or B star, the stellar wind can be very intense. Sometimes the compact object in a close orbit accretes material from the dense stellar wind and causes the weak X-ray emission [9].

are O or B stars with bright optical emission lines originated by circumstellar disc. It is believed that this envelope around the Be star is caused by its fast rotation. The orbit of the NS around the Be star is usually eccentric, so when is far away from the companion star cannot accrete material from this envelope. When the NS approaches periastron, it will be able to accrete material and the observer sees an X-ray outburst. Figure 2 shows a collapsed object in an eccentric orbit around a Be star (for a review of Be/X-ray binaries, [11, 12]).

However, a new kind of HMXBs was discovered by INTErnational Gamma Ray Astrophysical Laboratory (INTEGRAL) , called Supergiant Fast X-ray Transients [13], which are characterized by the occurrence of very fast X-ray outbursts. In [13] they have shown that at least a significant fraction of them are associated with supergiant stars.

Accreting X-ray pulsars were discovered by Giacconi et al. in 1971 [14] when they discovered the existence of periodic pulsations in the X-ray emission from Centaurus X-3. X-ray pulsars are rotating, highly magnetized NS accreting material from a companion binary star [15, 16, 17, 18]. Most known accreting X-ray pulsars belong

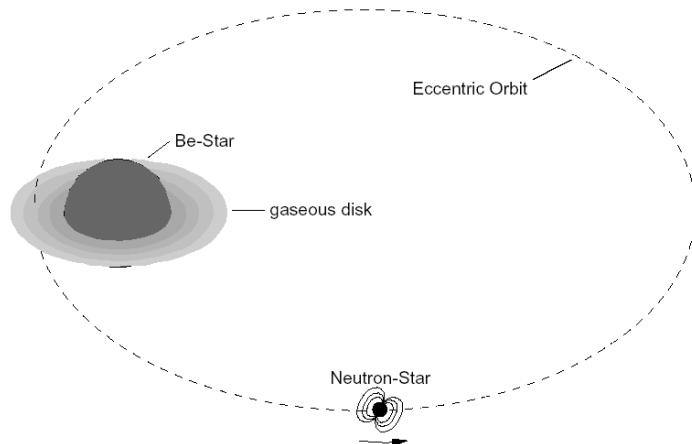


Figure 2: The NS revolves around the Be star in an eccentric orbit. When the NS comes inside the circumstellar disk of the Be star, accretion takes place and causes an X-ray outburst. However, while the NS is far away from the Be star, it is in quiescence and there is no X-ray emission.

to the HMXB class, such as 4U 1538–52/QV Nor. For accreting binary pulsars we measure their magnetic fields through the presence of cyclotron resonance scattering features (CRSFs) in their X-ray spectra because the cyclotron energy ( $E_{cyc}$ ) and the magnetic field strength ( $B$ ) are related to each other as  $E_{cyc} = 11.6 B(10^{12} G) (1+z)^{-1}$  keV, where  $z$  is the gravitational redshift.

## 2 Observational data

RXTE observed 4U 1538–52 between 1996 November 24 and 1997 December 13. To obtain the exact orbital phase we used the best fit orbital ephemeris from [19]. The exact time and the orbital phase are listed in [20].

In our analysis we used data from both RXTE pointing instruments, the Proportional Counter Array (PCA) and the High Energy X-ray Timing Experiment (HEXTE). To extract the spectra, we used the standard RXTE analysis software FTOOLS.

The PCA consists of five co-aligned Xenon proportional counter units with a total effective area of  $\sim 6000$  cm<sup>2</sup> and a nominal energy range from 2 keV to over 60 keV ([21]). However, due to response problems above  $\sim 20$  keV and the Xenon-K edge

around 30 keV, we restricted the use of the PCA to the energy range from 3 keV to 20 keV (see also [22]).

The HEXTE consists of two clusters of four NaI(Tl)/CsI(Na) Phoswich scintillation detectors with a total net detector area of 1600 cm<sup>2</sup>. These detectors are sensitive from 15 keV to 250 keV ([23]). However, response matrix, instrument background and source count rate, limit the energy range from 17 to 100 keV. Background subtraction in HEXTE is done by source-background swapping of the two clusters every 32 s throughout the observation. In order to improve the statistical significance of the data, we added the data of both HEXTE clusters and created an appropriate response matrix by using a 1:0.75 weighting to account for the loss of a detector in the second cluster. We also binned several channels together of the HEXTE data at higher energies and chose the binning as a compromise between increased statistical significance while retaining a reasonable energy resolution.

### 3 X-ray spectral analysis

The X-ray spectrum in accreting X-ray pulsars has been investigated during the last two decades. Nevertheless, there still exists no convincing theoretical model for the continuum of this kind of sources ([24], and references therein). Therefore, we have to use empirical models of the continuum in the fitting process. In the RXTE energy band these models take the general form of a power law times an exponential above a characteristic cutoff energy.

#### 3.1 Continuum model

We achieved a good description of the continuum X-ray spectra of this source using the standard pulsar continuum shape and customize models from XSPEC [20]. All of them are modified by a photoelectric absorption at low energies, a fluorescence iron emission line at 6.4 keV and the fundamental CRSF at 20 keV discovered by Ginga. The description of the continuum by physical models gave parameters that were not acceptable. We also modeled the observational data with several other standard pulsar continuum but they did not describe the spectra properly in the 7–16 keV energy band [25]. Therefore, in this paper we described the continuum produced in the accretion column of the NS by the Negative Positive power laws EXponential (NPEX) component modified by previous features. It was introduced by [26] and given by the formula:

$$NPEX(E) = A \cdot \left( E^{-\Gamma_1} + B \cdot E^{+\Gamma_2} \right) \cdot e^{-E/E_{fold}}, \quad (1)$$

where  $\Gamma_1$  and  $\Gamma_2$  are positive and  $E_{fold}$  is the folding energy of the high energy exponential cutoff. We used a Gaussian emission line at  $\sim 6.4$  keV due to the fluorescence

iron line and a multiplicative factor to fit the cyclotron line of the form:

$$cyclabs(E) = \exp\left(\frac{-\tau_c \cdot \left[\frac{\sigma_c E}{E_c}\right]^2}{(E - E_c)^2 + \sigma_c^2}\right), \quad (2)$$

where  $\tau_c$ ,  $\sigma_c$  and  $E_c$  are the optical depth, the width and the energy of the CRSF, respectively. We found that *cyclabs* model provides reasonable fits to the data. The observed X-ray spectrum of 4U 1538–52 is modified by photoelectric absorption due to the stellar wind of QV Nor.

### 3.2 CRSF variability

The main aim of this paper is to study the variation of the parameters of the cyclotron absorption line and their relations. The CRSF centered at  $\sim 20$  keV varies by  $\sim 15\%$  through the pulse [27]. In our pulse phase averaged spectra the parameter identified with the magnetic field,  $E_c$ , does not depend significantly on the orbital phase as is expected because it originates in the polar caps of the NS or the accretion column where the X-rays and the pulse forms [20]. In fact, the energy of the CRSF only varies by  $\sim 4\%$  (in some cases the same that the uncertainties of the parameter at 90% confidence level). Therefore, the stellar wind does not modify the accretion onto the NS during an orbital period significantly. However, we found that the relative width of the CRSF and its optical depth are correlated, as well as the CRSF energy and its width.

Although pulse phase resolved spectroscopy allows us to study the variation of the pulsar emission over the X-ray pulse, it seems that the correlations between the parameters of the CRSF are indeed real in pulse phase averaged spectroscopy. Using a consistent set of models [28] Coburn et al. parameterized a large sample of X-ray pulsars observed with RXTE which exhibit cyclotron lines. They found a new correlation between the relative width of the CRSF,  $\sigma_c/E_c$ , and its optical depth,  $\tau_c$ . Although they used pulse phase averaged spectra, another study of GX 301–2 [29] with pulse phase resolved spectra confirmed this correlation. They also reported another correlation between the cyclotron line width and the energy of the cyclotron line, both in pulse phase average and resolved spectra. Therefore we have plotted our fitted results for the cyclotron line parameters to check these correlations.

In Figure 3 we show the first correlation between the CRSF parameters in the  $\sigma_c - E_c$  plot. For a self-emitting atmosphere, the cyclotron line width and his energy is given by [30]:

$$\sigma_c \approx E_c \left(8 \ln 2 \frac{k T_e}{m_e c^2}\right)^{\frac{1}{2}} |\cos \theta|, \quad (3)$$

where  $k T_e$  is the temperature of the electrons along the magnetic field lines,  $E_c$  the cyclotron line energy, and  $\theta$  the viewing angle with respect to the magnetic field.

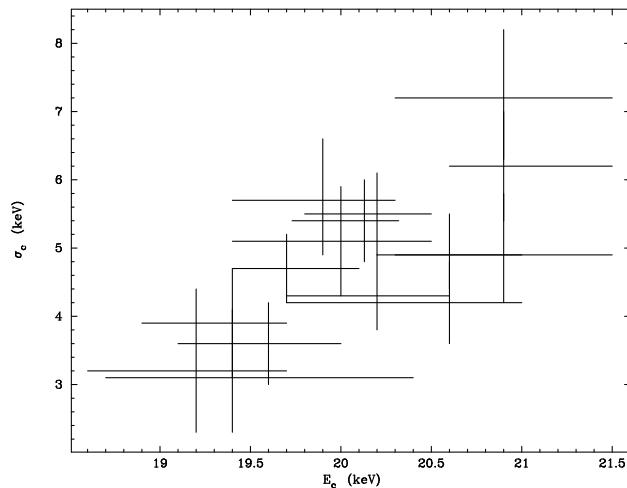


Figure 3: CRSF width  $\sigma_c$  versus CRSF energy  $E_c$  in 4U 1538–52. Note the moderate correlation between them. The bars indicate the uncertainties at 90% confidence level.

Therefore a linear correlation is only possible if  $\cos\theta$  does not change significantly. Our results imply an angle close to zero when  $k T_e$  is 5.2 keV and can take a value of  $18^\circ$  if the energy of electrons is 5.8 keV.

The parameter  $E_{fold}$  in the NPEX model is the typical temperature of the X-ray emitting plasma in keV. Basic Comptonization theory suggests that  $k T_e$  can be estimated from the folding energy of the pulsar continuum. Furthermore, if we assume that the seed photons for the Compton scattering in the accretion column are created throughout the volume of the accretion column, then detailed Monte Carlo simulations show that the optical depth of the CRSF is expected to be largest when the line of sight is almost perpendicular to the direction of the magnetic field [31]. If the temperature in the accretion column is constant, these models predict an anti correlation between the optical depth and the relative width of the CRSF. As we show in Figure 4, our phase averaged spectra results for 4U 1538–52 indicate a moderate correlation opposite to the models, and the conclusion is that the temperature in the accretion column is not constant. This correlation indicates that as CRSF increase in optical depth, the CRSF relative width increases as well.

In the above discussion, it has been assumed that the X-ray emission is produced from one homogeneous emission region. However, this source has contributions from both magnetic polar caps, which could influence the observed correlation. In fact, in 4U 1538–52 the two polar caps are observed and are unequal and non antipodal [32].

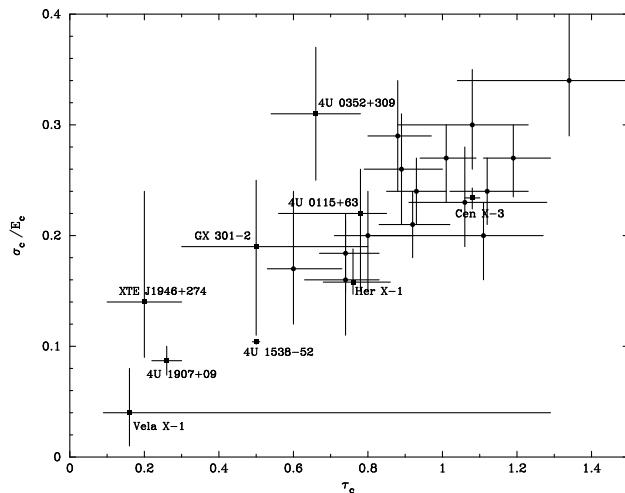


Figure 4: Fractional CRSF width  $\sigma_c/E_c$  versus the optical depth of the CRSF for several accreting pulsars from RXTE. Filled squares: values derived by [28] from phase averaged spectra. Filled circles: values derived from phase averaged spectra for 4U 1538–52 (this work, the bars indicate the uncertainties at 90% confidence level). All quantities refer to the pulse phase averaged values. Note the moderate correlation between them.

Therefore we can expect the parameters of the X-ray continuum emitted by each pole to be different, which can be reflected by changes in the observed continuum parameters. The variation in the folding energy of this system could explain this correlation caused by a mixture flux from the two polar caps. However, this reason cannot be the only one because, for example, in GX 301–2 the folding energy does not change significantly [29].

Assuming that equation 3 is correct and  $\cos\theta$  does not change appreciably, we expect a relationship between the folding energy and the relative width of the CRSF. As we can see in Figure 5, we found that the relationship was consistent with a power law  $\sigma_c/E_c \propto E_{fold}^{0.5}$ , indicating little variation of the angle  $\theta$  and according to equation 3. However, the statistic of the data is such that, from Figure 5, we cannot distinguish between the last relationship and a linear correlation  $\sigma_c/E_c \propto E_{fold}$ , at a statistically significant level.

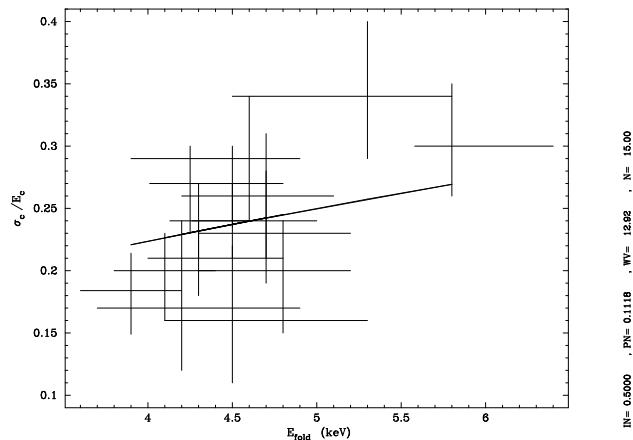


Figure 5: Fractional CRSF width  $\sigma_c/E_c$  versus the folding energy. All quantities refer to the pulse phase averaged values. The bars indicate the uncertainties at 90% confidence level.

## 4 Summary and conclusions

We have studied the variability of the parameters of the fundamental cyclotron absorption line in the HMXB 4U 1538–52. In order to explain this variability, we have also studied the relationship among the continuum ( $E_{fold}$ ) and cyclotron line parameters.

As shown in Figure 4, our pulse phase averaged results for 4U 1538–52 are in agreement with the correlation found by [28]. Also the pulse phase resolved results for GX 301–2 [29] noticed the same correlation, so it suggests that it is not due to effects of averaging. In terms of the relativistic cross sections, this result is in the opposite sense.

Furthermore, the fundamental CRSF of this source has a width nearly proportional to its energy. If we consider only a thermal broadening, then the variation of the viewing angle is not significant.

Our third result is a relationship between the folding energy  $E_{fold} = k T_e$  and the relative width of the cyclotron line. Figure 5 shows this correlation and it implies viewing angle close to zero which is consistent with the previous correlation.

Our results for the spectral analysis of the RXTE data of 4U 1538–52 can be summarized as follows:

- The absorbed NPEX continuum model provides a reasonable description of all the spectra used in this research. It approximates a photon number spectrum

for an unsaturated thermal Comptonization in a plasma of temperature  $T_e$ .

- The correlation between the width of the fundamental CRSF and its energy suggests little variations of the angle between the line of sight and the magnetic field in the accretion column.
- The relationship between the folding energy and the relative width of the fundamental CRSF implies that  $\cos\theta$  is close to 1. As an example,  $\cos 26^\circ \sim 0.9$ , so a little variation of the viewing angle and its relationship is consistent.
- The correlation between the relative width of the cyclotron line and its optical depth implies changes in the angle between the line of sight and the magnetic field at the NS poles, but in the opposite sense to the numerical simulations for the optical depth of the line.

*Acknowledgements:* This research is partially supported by the Spanish *INTEGRAL: Observaciones multifrecuencia de sistemas binarios de rayos X* project number ESP2001-4541-PE and *International Gamma Ray Astrophysics Lab. Operaciones. C3* project number ESP2002-04124-C03-03. This research has made use of data obtained through the High Energy Astrophysics Science Archive Research Center Online Service, provided by the NASA/Goddard Space Flight Center.

## References

- [1] Pfahl, E., Rappaport, S., Podsiadlowski, P. 2003, ApJ 597, 1036
- [2] Lewin, W.H.G., van Paradijs, J. and van den Heuvel, E.P.J. 1995, “X-ray binaries”, Cambridge University Press
- [3] Longair, M.S. 2004, “High Energy Astrophysics (Volume 2)”, Cambridge University Press
- [4] Frank, J., King, A., Raine, D. 2002, “Accretion Power in Astrophysics (3rd Edition)”, Cambridge University Press
- [5] Lewin, W.H.G., van der Klis, M. 2006, “Compact Stellar X-ray sources, Cambridge Astrophysics Series 39”, Cambridge University Press
- [6] Holt, S.S., McCray, R. 1982, Ann. Rev. Astron. Astrophys. 20, 323
- [7] Cominsky, L. 2001, “Procs. of the XXVI SLAC Summer Institute on Particle Physics, Gravity: from the Hubble length to the Planck length”, Stanford Linear Accelerator Center R-538 edited by Lance Dixon
- [8] Giovanelli, F., Sabau-Graziati, L. 2003, Chi. J. Astron. Astrophys. Vol. 3, 202
- [9] Shakura, N.I. and Sunyaev, R.A. 1973, A&A 24, 337
- [10] Liu, Q.Z., van Paradijs, J., van den Heuvel, E.P.J. 2000, A&AS 147, 25
- [11] Negueruela, I. 1998, A&A 338, 505

- [12] Ziolkowski, J. 2002, *Mem. S. A. It.* 73, 1038
- [13] Negueruela, I., Smith, D.M., Reig, P. et al. 2006, “*Procs. of The X-ray Universe 2005*”, ESA SP-604, 165
- [14] Giacconi, R., Gursky, H., Kellog, E. et al. 1971, *ApJ* 167, L67
- [15] White, N.E., Swank, J.H., Holt, S.S. 1983, *ApJ* 270, 711
- [16] Nagase, F. 1989, *PASJ* 41, 1
- [17] Bildsten, L., Chakrabarty, D., Chiu, J. et al. 1997, *ApJSS* 113, 367
- [18] Staubert, R. 2003, *Chi. J. Astron. Astrophys.* Vol. 3, 270
- [19] Makishima, K., Koyama, K., Hayakawa, S., Nagase, F. 1987, *ApJ* 314, 619
- [20] Rodes, J.J. 2007, Ph. D. thesis, University of Alicante
- [21] Jahoda, K., Swank, J.H., Giles, A.B. et al. 1996, “*EUV, X-ray and Gamma-Ray Instrumentation for Astronomy VII Proc. SPIE*”, ed. O. H. Siegmund and M. A. Gummin, vol. 2808, SPIE, p. 59
- [22] Kreykenbohm, I., Coburn, W., Wilms, J. et al. 2002, *A&A* 395, 129
- [23] Rothschild, R.E., Blanco, P.R., Gruber, D.E. et al. 1998, *ApJ* 496, 538
- [24] Harding, A.K. 1994, *AIP Conf. Proc.* 308, “*The Evolution of X-ray Binaries*”, ed. S. Holt & C. S. Day, p. 429
- [25] Rodes, J.J., Torrejón, J.M., Bernabéu, G. 2006, “*Proc. The X-ray Universe 2005*”, ESA SP-604, 287
- [26] Mihara, T. 1995, Ph. D. thesis, University of Tokyo
- [27] Clark, G.W., Woo, J.W., Nagase, F. et al. 1990, *ApJ* 353, 274
- [28] Coburn, W., Heindl, W.A., Rothschild, R.E. et al. 2002, *ApJ* 580, 394
- [29] Kreykenbohm, I. 2004, Ph. D. thesis, University of Tübingen
- [30] Mészáros, P., Nagel, W. 1985, *ApJ* 298, 147
- [31] Isenberg, M., Lamb, D.Q., Wang, J.C.L., 1998, *ApJ* 505, 688
- [32] Bulik, T., Riffert, H., Mészáros, P. et al. 1995, *ApJ* 444, 405



# IR PROPERTIES OF ASTROPHYSICAL ICES IN THE UNIVERSE

ROSARIO VILAPLANA<sup>1</sup>, OSCAR GOMIS<sup>1</sup>, MIGUEL A. SATORRE<sup>1</sup>, RAMON LUNA<sup>1</sup>, JOSE CANTO<sup>1</sup>, MANUEL DOMINGO<sup>1</sup>, CARLOS MILLAN<sup>1</sup>, CARMINA SANTONJA<sup>1</sup>

*1 Departamento de Física Aplicada, Escuela Politécnica Superior de Alcoy, Universidad Politécnica de Valencia, E-03801 Alcoy, SPAIN*

**Abstract:** Part of our knowledge about interstellar, planetary and cometary ices derives from the identification of bands in astronomical spectra of distant objects, rather than from direct measurement. Therefore, laboratory experiments on simple ices and its mixtures carried out under physical conditions as close as possible to the astrophysical environments are essential for the identification of ices present in space.

In this work, we first give the reader an introduction to the infrared properties of the ices present in the Universe, defining the term ice and showing the importance of infrared spectroscopy for the study of ices, the locations in space where ices are mainly found along with the ices detected so far, and some of the experimental laboratories around the world working on ice studies.

The second part of this work is focused on the Experimental Astrophysics Laboratory we have set up in order to characterize in the mid and far infrared range (MIR and FIR respectively), which covers the 2.5-1000  $\mu\text{m}$  region, ices of astrophysical interest. One of the key components of the laboratory is a Si composite bolometer cooled by liquid He used to obtain spectra with a sensitivity much greater than that obtained with a standard DTGS detector working at room temperature in the FIR where bands of small intensity lie. The MIR and FIR spectra we shall obtain in the future will serve to characterize in those spectral ranges some simple ices and their mixtures and also to compare data mainly coming from the forthcoming Herschel satellite with laboratory data. It is also planned to do a systematic laboratory study of the effects that ultraviolet (UV) photolysis and temperature have on the ice band profiles and on the ice structure.

**Keywords:** Ices – Infrared Spectroscopy – Laboratory Experiments – Interstellar Medium – Solar System.

## 1 Introduction

Astronomy studies the properties of the radiation (light) reaching us from celestial objects. Cool objects – up to temperatures of 3500 °C – radiate most of their energy at infrared (IR) wavelengths while hotter objects like the Sun radiate strongly at shorter wavelengths. The cool Universe is therefore best studied in the IR region. The Universe is full of cool objects, such as planets, dust and ageing stars, none of which generally shine brightly in the optical part of the spectrum. In fact, observations of these objects began with the arrival of sensitive IR detectors [1]. The first IR observations with a space-based telescope were carried out by the Infrared Astronomical Satellite (IRAS) [2]. Twelve years later, the Infrared Space Observatory (ISO) was launched [3]. Currently, the Spitzer Space Telescope [4] and the AKARI mission [5] are making an all-sky survey obtaining spectra of much better sensitivity than the previous ones. On the other hand, these missions are all limited to wavelengths below 200  $\mu\text{m}$ . Nevertheless, the upcoming Herschel Space Observatory (HERSCHEL) [6] will provide spectra of high spectral resolution from 55 to 672  $\mu\text{m}$  covering the far infrared and submillimetre spectral range.

Each substance produces a different spectrum which can therefore be used as an identifier. IR spectra are one of the tools used to extract great amount of information from objects in the Universe. Some bands (features) of these IR spectra are caused by species in the form of ices.

A good understanding of the term “ice” is needed. The term ice typically refers to a solid composed of molecules which form a gas or a liquid at 25 °C and 1 bar. A distinction between a one-component and a multi-component solid is also needed. With this purpose expressions such as “pure ice” and “mixed-molecular ice” can be employed, respectively [7].

Ices absorb specific wavelengths of light at IR frequencies depending on their molecular bonding and composition. The mid-infrared region (4000–400  $\text{cm}^{-1}$  or 2.5–25  $\mu\text{m}$ ) contains the infrared active fundamental bands (intramolecular vibrations) as well as some low order overtone and combination modes [8]. The broad absorptions in the far infrared region (400–10  $\text{cm}^{-1}$  or 25–1000  $\mu\text{m}$ ) are due to lattice modes (intermolecular vibrations) [9]. Moreover, ice bands can be distinguished from bands of gas phase molecules because ice bands lack the characteristic ro-vibrational structure of gases and are shifted and broader.

One of the scenarios where ices are thought to be present is on the surface of dust grains in the interstellar and circumstellar medium [10]. In fact, ices are generally one of the major components of the cold clouds in the interstellar medium and show large variations in abundances and profiles which means they can be used as powerful diagnostics of changes in astrophysical environments. Some bodies in the Solar System where ices have been reported are the surfaces of some planets and satellites, some small bodies – comet nuclei, Trans-Neptunian Objects (TNOs), centaurs– and even

some planetary rings [11] and some comet tails [12]. On the other hand, comets are considered to be the most primitive objects in the Solar System. The composition and structure of their nuclei contain a record of the primordial solar nebula at the time of its formation. Cometary nuclei are made of refractory solids and frozen volatiles. Moreover, the composition of volatiles is thought to be similar to that observed in dense molecular clouds. It therefore seems that there is a close relationship between cometary materials and interstellar icy grain mantles [7].

Dust grains are microscopic particles whose sizes are usually less than 1 micrometre. These grains are represented by a core of refractory materials (silicates, amorphous carbon, and organic residues) covered with an icy mantle largely dominated by H<sub>2</sub>O ice but also with other major components as CO, CO<sub>2</sub>, NH<sub>3</sub>, CH<sub>3</sub>OH, and H<sub>2</sub>CO ices. The visible light wavelength is of the same size of many dust particles, consequently the visible light is scattered by the dust; however, longer wavelengths such as IR radiation passes through. On the other hand, the dust itself is also a source of IR radiation that can be picked up by detectors. For instance, stars, depending on their temperature, emit in a particular range and it is possible to detect their absorbing spectra. Besides, this starlight may be absorbed by the surrounding dust grains which will be warmed up and start to radiate in the IR wavelength. This absorption of energetic radiation and reemission at less energetic wavelengths is very efficient and dust clouds emit the majority of their energy at IR wavelengths. A similar situation occurs with dust grains of cometary tails when they are approaching the Sun.

In the website of the Cosmic Ice Laboratory at NASA's Goddard Space Flight Center [13] we can find listed the Solar System ices and also the detected molecules in the interstellar medium and in comets. Some of the molecules of the interstellar medium have been identified in their ice phase as well as their gas phase (e.g. H<sub>2</sub>O, CO, CO<sub>2</sub>, OCS, NH<sub>3</sub>, CH<sub>4</sub>, CH<sub>3</sub>OH). On the other hand, the existence of a great number of molecules in gas phase—some of them with a high number of atoms (e.g. benzene)—is significant; in fact it indicates that a rich chemical activity is taking place inside these environments. There is also in this website a database containing MIR and FIR reference ice spectra. Other research groups have also carried out ice spectroscopy applied to different astrophysical environments, namely, the Leiden Molecular Astrophysics group in Holland [14], the Experimental Astrophysics Laboratory group at Catania in Italy [15], the NASA Ames Research Center group in USA [16], the Laboratory Astrophysics Facility group at University College in Australia [17], and the Laboratory for Atomic and Surface Physics at Virginia in USA [18]. Among all these groups, the Cosmic Ice Laboratory at NASA's Goddard Space Flight Center, the Laboratory Astrophysics Facility at University College, and the NASA Ames Research Center group have carried out ice spectroscopy in the FIR up to 100, 110, and 200  $\mu\text{m}$ , respectively.

The abundance of different ices depends on the astrophysical environment. H<sub>2</sub>O ice is clearly dominant in almost all these environments. In the case of Solar System

bodies where the ices are found mainly on the surfaces of the bodies, most of the identified surface ices consist primarily of water. Some exceptions are the surfaces of Triton and Pluto which are covered with carbon and nitrogen containing ices and the surface of Io which is dominated by sulphur containing ices. In the case of interstellar ices, in addition to the H<sub>2</sub>O ice, other major ice components are CO, NH<sub>3</sub>, CH<sub>3</sub>OH, CO<sub>2</sub> and H<sub>2</sub>CO. In the case of cometary nuclei the composition of the volatile ices is largely dominated by H<sub>2</sub>O ice (about 70–90%) while other major ice components are CO, CH<sub>3</sub>OH, CO<sub>2</sub>, and H<sub>2</sub>CO. The similarity between the volatile ices in comets and interstellar ices tends to support a close relationship between cometary materials and interstellar icy grain mantles. Thus, cometary ices constitute a link between interstellar and Solar System materials [19, 20, 21].

As said previously, ices in the interstellar and circumstellar medium are detected on silicatic and carbonaceous grain mantles. In the formation and evolution of those ices a number of different processes play a role. In fact, the ices can be formed by sublimation and recondensation after being synthesized on grain surfaces or directly frozen from gas phase as in the case of CO. However, other more energetic ways of ice synthesis are also possible such as UV photolysis, ion irradiation and thermal annealing. The Solar System and cometary ices are also subjected to radiation processing coming from the magnetospheres, cosmic rays, solar wind and flares, and solar UV field. The effect that energetic charged particle irradiation and UV photolysis have on the ices is that when an energetic ion or UV photon collides with the ice, part of the deposited energy destroys bonds in the target, forming radicals that can then react to synthesize new molecules. In this way, new molecules are produced by irradiation being an example of this phenomenon the hydrogen peroxide molecule which has been detected on the surface of the Galilean satellite Europa, whose surface is being bombarded by ions coming from the intense Jupiter magnetosphere [22]. Temperature variations in astrophysical scenarios may alter the ice phase producing a phase change via amorphization or crystallization. To study the role played by each of the latter processes in ice evolution, different groups are carrying out laboratory experiments simulating the astrophysical scenario under study. In this way, experiments such as UV photolysis [23, 24, 25, 26, 27], ion irradiation (radiolysis, e.g., [28, 29, 30, 31, 32, 33]) and surface chemistry [34, 35] provide us with information about the basic processes of ice formation, adsorption, desorption, diffusion, and even reaction pathways.

Lastly, laboratory experiments can shed some light on, for example, how the intrinsic strength of each band depends on the chemical environment of the molecule, or how the position of the stretching mode shifts by 20 cm<sup>-1</sup> and its width broadens by a factor of two when the CO<sub>2</sub> is mixed with H<sub>2</sub>O ice rather than in its pure form, or how the ice profile of crystalline material can be distinguished from their amorphous counterparts by being sharper and red-shifted and, finally, how to analyze the band profiles making studies with layered ices, in which some molecules (e.g., CO) are not

mixed with H<sub>2</sub>O [36]. In other words, laboratory experiments on simple ices and their mixtures carried out under physical conditions as close as possible to the astrophysical environments are essential to provide the basic data to analyze quantitatively the ice bands (e.g. [37, 38, 39, 40])

## 2 Description of the Experimental Astrophysics Laboratory

### 2.1 Experimental equipment

Our astrophysics laboratory located at the Escuela Politécnica Superior de Alcoy (Valencia, Spain) is composed of several basic parts. The main one is an experimental chamber that is a high vacuum device where a pressure of  $10^{-7}$  mbar is obtained using a turbomolecular pump in a High Vacuum System backed by a root pump. A compressor connected with a closed-cycle helium cryostat is used to cool the sample down to 10 K. A resistor and a temperature controller permit varying the temperature up to 300 K. This temperature is measured with two silicon diodes at different locations on the sample holder. The gas molecules or a mixture of them prepared in a pre-chamber pass through a needle valve to the chamber when the proper conditions are reached. The velocity of deposition is controlled and regulated by the aperture of this valve and the pressure in the pre-chamber. The proportion among the gases is controlled by their partial pressure measured with a ceramic sensor not influenced by the gas type. The sample holder is a Q-sense Quartz Crystal Microbalance (QCMB) whose frequency is measured precisely with a frequencymeter.

The second part is a mass spectrometer which is used to analyze the composition of the gases. This spectrometer is connected at the bottom of the vacuum chamber getting a full scan of all the species present in the chamber. The result is registered as the percentage in volume of every gas present. The spectrometer also offers the option of selecting specific species to be scanned being the measured and registered magnitude the partial pressure. Because the spectrometer takes pressure values every 4 or 5 seconds, it is possible to study the temporal evolution of these pressures.

The third part is an ultraviolet Hydrogen lamp that provides ultraviolet (UV) photons, which are used to irradiate the ice sample. The photon flux should be about  $10^{15}$  photons  $\text{cm}^{-2} \text{s}^{-1}$  with energy  $E_{\text{photons}} \geq 6$  eV and its spectrum is dominated by five bands. We usually irradiate ice films with a thickness of 0.1–0.2  $\mu\text{m}$  to be sure that the sample is optically thin. This irradiation destroys and produces new species which can be detected in different ways. However, they are very well observed with IR spectroscopy.

The fourth part is the IR spectrometer, a Fast Fourier Bruker IFS 66v/S, which integrates a Michelson interferometer obtaining spectra with a resolution better than

$0.25\text{ cm}^{-1}$ . The spectrometer operates under vacuum conditions of  $10^{-1}$  mbar in the optics and the sample compartment. This environment makes it possible to record spectra free from gas phase interference such as  $\text{H}_2\text{O}$  or  $\text{CO}_2$  in the MIR and especially for measurements in the FIR where water vapor contamination is very pronounced. The basic components of a spectrometer are the source, the beamsplitter and the detector. These components must be selected depending on the spectral range under study; in our particular case the MIR and the FIR. We are equipped with the Globar and the Hg-Arc sources, for the MIR and FIR, respectively and with a KBr beamsplitter for the MIR and four Mylar beamsplitters (6, 25, 50 and  $100\ \mu\text{m}$ ) covering all the FIR range. Concerning the detectors we have two DTGS (deuterated triglycine sulfate) for the MIR and the FIR, respectively, and a Si bolometer -a very high sensitivity detector- covering the spectral range  $670\text{--}10\text{ cm}^{-1}$  ( $15\text{--}1000\ \mu\text{m}$ ). The two greatest disadvantages of the DTGS detectors are their low sensitivity and their slowness. On the other hand, the Si bolometer allows measurements of high sensitivity in much shorter time. Nevertheless, this detector must be cooled with liquid helium and its dewar pre-cooled and refrigerated with liquid nitrogen being its use quite complex. Moreover, liquid helium is very expensive and evaporates quickly which implies that the experiments must be well planned.



Figure 1: The spectrometer, the 3D movable high vacuum chamber, the silicon bolometer and the helium container

Figure 1 is showing the experimental configuration designed by us in order to obtain IR spectra of simple ices and its mixtures in the MIR and FIR. This configuration integrates the IR spectrometer, the high vacuum chamber and the Si bolometer

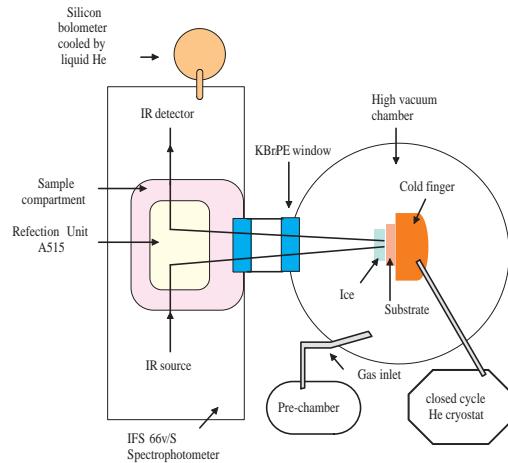


Figure 2: Block diagram showing the IR spectrometer, the bolometer and the high vacuum chamber

together. In order to offer the possibility of carrying out several experiments on the same sample at the same time the high vacuum chamber is located outside the sample compartment of the IR spectrometer making the four chamber's windows available. An essential part of this design is a mechanism that allows us to move the high vacuum chamber in the three spatial dimensions by using a micrometric adjustment, which allows the right alignment of the IR beam when it travels from the IR source to the detector.

Figure 2 shows a bottom view of our design, the IR spectrometer on the left and the high vacuum chamber on the right. The technique we are using to obtain measurements is the TRT (Transmission-Reflection-Transmission). The beam is directed to the cold finger in the chamber by using the reflection module A515. There is a vacuum of  $10^{-1}$  and of  $10^{-7}$  mbar inside the interferometer sample compartment and the ice chamber, respectively. In order to solve the problem with the vacuum, when the beam is traveling outside and between the interferometer and the ice chamber, we use nitrogen gas. As the high vacuum chamber is located outside the interferometer sample compartment forces it to work in TRT. However, this fact allows us to carry out different experiments at the same time on the same ice sample. We are considering this possibility because it can give complementary information about the same ice sample. A modification of the high vacuum chamber is being done; specifically two quartz crystal microbalances covered by their respective plane optically thick gold films have been placed on both sides of the cold finger in a way that simultaneous

gas depositions take place.

Finally, in our laboratory we also use a UV-Visible spectrometer EPP-C 2000 from Stellar.net.

## 2.2 Measurements carried out

Bellow, we describe the experiments we are carrying out at the present time. We have measured the density and the index of refraction real part of simple ices (e.g. CH<sub>4</sub>, N<sub>2</sub> and CO<sub>2</sub>) and their mixtures at low pressure ( $P < 10^{-7}$  mbar) and at temperatures ranging from 10 K up to the sublimation temperature of the species or mixture under study [41]. The technique used has been double laser interferometry. There is a lack of density measurements for ices other than water at the low temperatures and pressures found in space environments and cryogenic laboratory systems. Moreover, the density is related with the porosity which is another useful parameter to characterize the ice structure. By knowing how the ice density varies with temperature some clues can be obtained about structural changes as a function of temperature. In order to obtain the real part of the refractive index in the UV-Visible range (200–800 nm), we use a UV-Visible lamp which is directed to the growing ice. The reflected beam is collected by a UV-Visible spectrometer and we can draw an interference curve each 0.2 nm in this range, that allows the calculation of the refractive index value [42].

We have also conducted some experiments on temperature desorption at a constant warming up rate. In particular, the ability of CO<sub>2</sub> ices to trap other gases such as CH<sub>4</sub> [43] has been studied. These kinds of studies are necessary to better understand the processes that occur in the surfaces of some Solar System bodies as in the case of Triton where ices of N<sub>2</sub>, CH<sub>4</sub>, CO, CO<sub>2</sub>, and H<sub>2</sub>O are present on its surface. In particular, the process of ice trapping in some ice matrix explains the fact that the trapped ice sublimates at higher temperature than usual.

Some experiments on UV irradiation of simple ices and its mixtures have also been carried out [44]. We have verified that CO<sub>2</sub> ice irradiation produces CO ice. As previously stated, UV irradiation produces new molecules originally not present in the target. At this moment, these new molecules have been detected by mass spectrometry although we have planned in the near future to use the IR spectrophotometer to characterize the new species by means of their IR absorption bands.

## 2.3 Measurements planned in the future

One of the aims of our laboratory in the forthcoming years will be to obtain high sensitivity spectra of molecular ices and their mixtures in the MIR and the FIR up to 670  $\mu\text{m}$  [45]. We plan to obtain the ice band profiles, peak positions, and FWHM as a function of temperature. By depositing the ices at different temperatures the

amorphous and crystalline phases of the ices will be obtained and characterized. Phase change studies will be another type of experiments that are scheduled.

In addition, with the design presented above we will focus on carrying out several experiments at the same time on the same sample by using techniques such as IR spectroscopy, UV irradiation, temperature desorption, UV-Visible spectroscopy, etc. [45].

### 3 Conclusions

In this work, we have presented an introduction to the infrared properties of ices in the Universe. The places in the Universe where ices are mainly found have been described both in the Solar System and in the interstellar medium. In addition, the astrophysical processes which determine the evolution of the ices (temperature changes, ionizing irradiation, adsorption and desorption processes, etc.) have been mentioned. At the Department of Applied Physics of The University of Valencia at Alcoi (Spain), we have set up an Experimental Astrophysics Laboratory to simulate the formation and subsequent processing of ices in space. Our main goal is to characterize simple ices and their mixtures. Currently, we are able to carry out experiments on UV irradiation, temperature desorption, UV-Visible spectroscopy, density measurements, and to obtain the index of refraction real part of ices and their mixtures. In the future, we plan to obtain IR spectra in the 2.5-670  $\mu\text{m}$  range of simple ices and their mixtures of astrophysical interest in order to compare laboratory data with spatial data coming from the forthcoming Herschel mission.

*Acknowledgements:* This work has been carried out under AYA2004-05382 project from the Spanish Ministerio de Educación y Ciencia, being partially supported by FEDER funds.

### References

- [1] Battrick, B. (Ed.) 2001, in “The Infrared Revolution”, ESA Brochure 170, ESA Publications Division
- [2] <http://irsa.ipac.caltech.edu/IRASdocs/iras.html>
- [3] <http://www.iso.vilspa.esa.es/>
- [4] <http://www.spitzer.caltech.edu/spitzer/index.shtml>
- [5] [http://www.ir.isas.jaxa.jp/ASTRO-F/Outreach/index\\_e.html](http://www.ir.isas.jaxa.jp/ASTRO-F/Outreach/index_e.html)
- [6] <http://www.rssd.esa.int/herschel/>
- [7] Hudson, R.L., Moore, M.H. 2001, JGR 106, 33275
- [8] Schmitt, B., Quirico, E., Trotta, F. 1998, in “Solar System Ices”, Kluwer Academic Press, 199

- [9] Moore, M.H., Hudson, R.L. 1994, *A&A Sup. Ser.* 103, 45
- [10] Ehrenfreund, P., Schutte, W. A. 2000, "Procs. of the Astro-Soc of the Pac.", 135
- [11] Schmitt, B., de Bergh, C., Festou, M. (Eds.) 1998, in "Solar System Ices", Kluwer Academic Press.
- [12] Lellouch, E., Crovisier, J., Lim, T. et al. 1998, *A&A* 339, 9
- [13] <http://www-691.gsfc.nasa.gov/cosmic.ice.lab/>
- [14] <http://www.strw.leidenuniv.nl/astrochem/>
- [15] <http://web.ct.astro.it/weblab/>
- [16] <http://www.astrochem.org/index.htm>
- [17] Hyland, A.R., Smith, R.G., Robinson, G. 1992, "Procs. of the Astronomical Society of Australia", vol. 10, 77
- [18] <http://www.virginia.edu/ep/LASP/LASP.html>
- [19] Ehrenfreund, P., Irvine, W., Becker, L. et al. 2002, "Procs. of the First European Workshop on Exo-Astrobiology", 9
- [20] Ehrenfreund, P., Fraser, H. 2003, "Procs. of the NATO Advanced Study Institute on Solid State Astrochemistry", vol. 120, 317
- [21] Crovisier, J. 2006, *Molecular Physics* 104, 2737
- [22] Carlson, R.W., Anderson, M.S., Johnson, R.E. et al. 1999, *Science* 283, 2062
- [23] d'Hendecourt, L.B., Allamandola, L.J., Grim, L.J.A. et al. 1986, *A&A* 158, 119
- [24] Allamandola, L.J., Sandford, S.A., Valero, G.J. 1988, *Icarus* 76, 225
- [25] Kouchi, A., Kuroda, T. 1990, *Nature* 344, 134
- [26] Gerakines, P.A., Schutte, W.A., Ehrenfreund, P. 1996, *A&A* 312, 289
- [27] Leto, G., Baratta, G. 2003, *A&A* 397, 7
- [28] Strazzulla, G., Baratta, G.A. 1992, *A&A* 266, 434
- [29] Moore, M.H., Hudson, R.L. 1992, *ApJ* 401, 353
- [30] Baratta, G.A., Castorina, A.C., Leto, G. et al. 1994, *Planet. Space Sci.* 42, 759
- [31] Brucato, J.R., Palumbo, M.E., Strazzulla, G. 1997, *Icarus* 125, 135
- [32] Moore, M.H., Hudson, R.L. 2000, *Icarus*, 145, 282
- [33] Strazzulla, G., Leto, G., Gomis, O. et al. 2003, *Icarus* 164, 163
- [34] Kimmel, G.A., Stevenson, K.P., Dohnalek, Z. et al. 2001, *J. Chem. Phys.* 114, 5284
- [35] Collings, M.P., Dever, J.W., Fraser, H.J. et al. 2003, *ApJ* 583, 1058
- [36] van Broekhuizen, F.A., Groot, I.M.N., Fraser, H.J. et al. 2006, *A&A* 451, 723
- [37] Hudgins, D.M., Sandford, S.A., Allamandola, L.J. et al. 1993, *ApJ Sup. Ser.* 86, 713
- [38] Gerakines, P.A., Schutte, W.A., Greenberg, J.M. 1995, *A&A* 296, 810
- [39] Ehrenfreund, P., Boogert, A.C.A., Gerakines, P.A. et al. 1996, *A&A* 315, L341
- [40] Ehrenfreund, P., Boogert, A.C.A., Gerakines, P. A. et al. 1997, *A&A* 328, 649
- [41] Vilaplana, R., Domingo, M., Satorre, M.A. et al. 2007, "Procs. of the 4th Annual meeting of the AOGS", 217
- [42] Millan, C., Domingo, M., Canto, J. et al. 2007, "Procs. of the Space Astronomy: The UV Window to the Universe", 81

- [43] Luna, R., Satorre, M.A., Millan, C. et al. 2007, "International Astrophysics and Astrochemistry Meeting: Molecules in Space & Laboratory", 122
- [44] Canto, J., Satorre, M.A., Luna, R. et al. 2006, "6th European Workshop on Astrobiology", 168
- [45] Vilaplana, R., Gomis, O., Luna, R. et al. 2007, "Procs. of the 4th Annual meeting of the AOGS", 198



# AN OVERVIEW OF SOHO 12 YEARS LATER

RAÚL GÓMEZ-HERRERO<sup>1,2</sup>, DOLORES RODRÍGUEZ-FRÍAS<sup>1</sup>, GEMA  
HEBRERO<sup>1</sup>, RICARDO CRUZ<sup>1</sup>, and LUIS DEL PERAL<sup>1</sup>, REINHOLD  
MÜLLER-MELLIN<sup>2</sup>, BERN HEBER<sup>2</sup>, R. WIMMER-SCHWEINGRUBER<sup>2</sup>  
*1 Grupo de Astropartículas y Plasmas Espaciales, Universidad de Alcalá, E-28871  
Alcalá de Henares, Madrid, SPAIN*  
*2 Institut für Experimentelle und Angewandte Physik, Universität Kiel, Kiel,  
GERMANY*

**Abstract:** The Solar and Heliospheric Observatory SOHO was launched on 1995 December 2 and completed launch and early orbit activities (commissioning) in 1996 March, reaching its L1 halo orbit. The control of the spacecraft was lost in 1998 June, and only restored three months later through the heroic efforts of an ESA-NASA-contractor-university team. All twelve scientific instruments were still usable, most with no ill effects. Despite the immediate failure of two of the three onboard gyroscopes and the later demise of the third (in 1998 December), by 1999 February, new, gyroless, onboard control software not only allowed the spacecraft to return to full scientific usefulness, but actually provided a greater margin of safety for spacecraft operations. SOHO's current margins for thruster propellant and solar panel output are considered by spacecraft engineers to be adequate to extend its operational life longer than 2010. Some of the scientific insights from SOHO and scientific topics for future are reviewed in the following lines.

**Keywords:** Solar Energetic Particles – SOHO satellite – Solar Physics – Solar Terrestrial Physics.

## 1 Introduction

The Solar and Heliospheric Observatory (SOHO) celebrates its 12th launch anniversary on December 2. In late 1996, shortly after its launch, SOHO was able to observe the last minimum of the roughly 11-year activity cycle of the Sun. The minimum was followed by a rapid rise in solar activity, peaking 2001 and 2002. Activity levels have slowly declined since then, but we haven't reached solar minimum yet, despite passing 11.1 years since the last minimum (the average length of a solar cycle, Figure 1).

SOHO, the Solar and Heliospheric Observatory, is an international cooperative project by ESA and NASA to study the Sun, from its deep core to the outer corona,

and the solar wind. It carries a complement of twelve sophisticated instruments developed and furnished by 39 institutes from fifteen countries (Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Japan, The Netherlands, Norway, Russia, Spain, Switzerland, United Kingdom, and the United States).

Primary scientific objectives of SOHO satellite are:

- Understand the changing flow of energy and matter throughout the Sun, heliosphere, and planetary environments;
- explore the fundamental properties of plasma systems; and
- define the origins and impacts of variability in the Sun-Earth system.

SOHO science is rooted in all three of these objectives.

Solar Interior are sonded by SOHO with heliosismology instrumentation:

- GOLF: Global Oscillations at Low Frequencies;
- VIRGO: Variability of Solar Irradiance and Gravity Oscillations; and
- MDI: Michelson Doppler Imager.

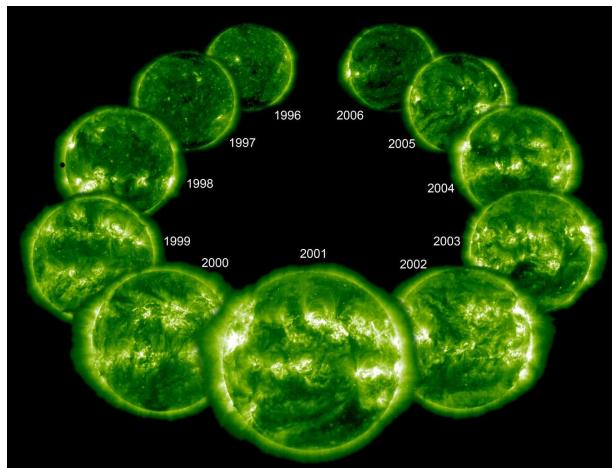


Figure 1: An EIT image from each year of nearly an entire solar cycle assembled by Steele Hill (NASA GSFC).

Solar atmosphere (Photosphere, Chromosphere and Corona) are directly observed in several region of the electromagnetic spectrum with:

- SUMER: Solar Ultraviolet Measurements of Emitted Radiation;
- CDS: Coronal Diagnostic Spectrometer;
- EIT: Extreme-Ultraviolet Imaging Telescope;
- UVCS: Ultra-Violet Coronagraph Spectrometer; and
- LASCO: Large-Angle Spectroscopic Coronagraph.

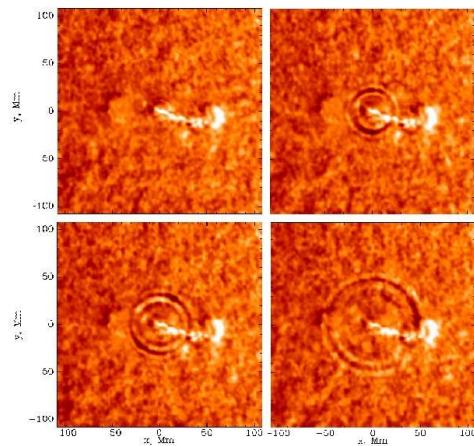


Figure 2: Scientists have shown for the first time that solar flares produce seismic waves in the Sun's interior that closely resemble those created by earthquakes on our planet. The researchers observed a flare-generated solar quake that contained about 40,000 times the energy released in the great earthquake that devastated San Francisco in 1906

Finally, solar wind and solar energetic particles are analysed with particle instrumentation on board SOHO:

- CELIAS: Charge, Element and Isotope Analysis System;
- SWAN: Solar Wind Anisotropies;
- COSTEP: Comprehensive Supra-Thermal and Energetic-Particle Analyser, with two instruments: LION (Low Energy Ion and Electron Instrument) and EPHIN (Electron, Proton Helium Instrument); and

- ERNE Energetic and Relativistic Nuclei and Electron Experiment also with two instruments: LED (Low Energy Detector) and HED (High Energy Detector).

SOHO has improved our understanding of the Sun since its launch. It has provided the first images of structures and flows below the surface of the Sun and of activity on the far side. SOHO has revealed the Sun's extremely dynamic atmosphere, provided evidence for the transfer of magnetic energy from the surface to the outer solar atmosphere, the corona, through a magnetic carpet, and identified the source regions of the fast solar wind. It has revolutionised our understanding of solar-terrestrial relations and dramatically improved our space weather-forecasting by its continuous stream of images covering the atmosphere, extended corona and far side.

SOHO findings are documented in 2500 papers since its launch, representing the work of 2300 scientists, approximately. At the same time, SOHO's easily accessible, spectacular data and fundamental scientific results have captured the imagination of the space science community and the general public alike.

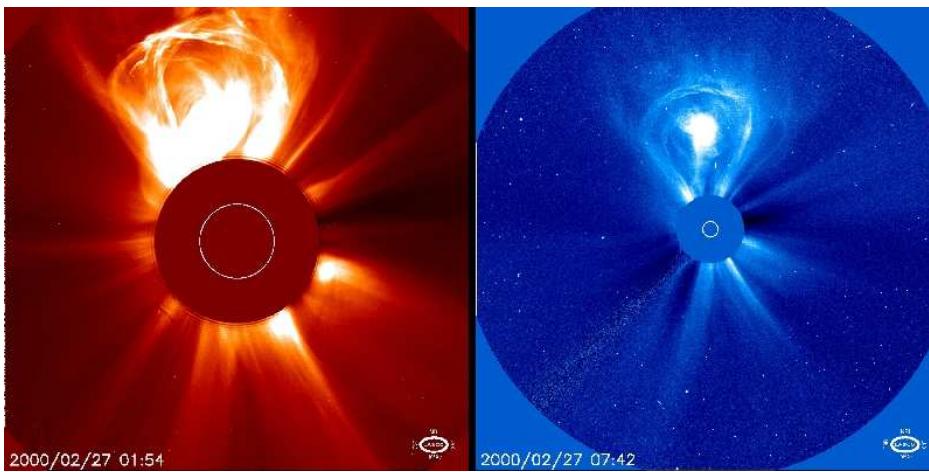


Figure 3: A coronal mass ejection on Feb. 27, 2000 taken by LASCO C2 and C3. A CME blasts into space a billion tons of particles travelling millions of miles an hour. This particular CME is “lightbulb-shaped”.

## 2 Scientific insights from SOHO

Among the scientific insights obtained from SOHO we can briefly enumerate a lot of advances in the following topics:

## 2.1 The Solar Interior and Total Irradiance

**Total solar irradiance variations.** VIRGO radiometers takes the dose of radiation over 12 years including a complete solar cycle. The total solar irradiance (TSI) obtained with the VIRGO experiment [1] makes possible the first independent and internally consistent determination of possible long-term changes, such as degradation. Interpretation of the TSI record, whether as a steady cycle with no underlying secular change or as showing an increasing trend, has broad social and political impacts as governments make decisions on their responses (if any) to global warming.

**Constraints on solar abundances.** The recent measurements on abundances of heavy elements [2], [3] have led to significant changes in our understanding of the internal structure and constitution of the Sun. Standard solar models calculated with the new abundances are significantly different from the helioseismic results which were in a good agreement with the previous solar models [2]. This discrepancy affects the predicted neutrino fluxes [4], [5].

**Rotation of the deep interior.** Analyzing SOHO GOLF data in this way, [6] obtain results that are consistent with a flat solid-body rotation rate of the inner layers down to around  $0.2 R_{Sun}$ . No g-modes have yet been positively identified, although some candidates have been found in the GOLF observations.

**Solar-cycle variations in the size of the Sun.** The MDI data analysis made that Dziembowski and Goode [7] concluded that the Sun becomes smaller and cooler at solar maximum. By contrast, it shows an increasing total irradiance with increasing solar activity from VIRGO data. MDI data have also revealed dramatic changes with the solar cycle of the large-scale, subsurface dynamics of the Sun [8], [9], [10].

**Sunquakes.** Sunquakes can provide new information about energy release and transport in solar flares. The first sunquake observed related to a flare in 1996 July 9. Circularly expanding waves on the solar surface were observed (Figure 2), the last X-class flare of the previous solar activity cycle and the first flare observed by SOHO. No sunquakes, however, were observed by SOHO until the X17 flare of 2003 October 28. This flare and two somewhat smaller flares, produce seismic waves observable with MDI. These new observations from SOHO provide unique information about the interaction of the high-energy particles associated with solar flares with the dynamics of the solar atmosphere during and after the impulsive phase of the flares, providing the new methods of helioseismologic analysis of flaring active regions, similar to that used in earthquakes study.

## 2.2 The Solar Atmosphere

Nagata et al. [11] found that hot loops visible in soft X-rays were distinct from the cooler loops visible in the EUV, and that, in general, the two types of loops alternated spatially. Neither of them seems to be isothermal.

Magnetic field as high as 1750 G at 6 Mm and 960 G at 12 Mm above a pair of large sunspots have been observed with MDI, CDS and EIT [12]. This means the first observation of plasma at temperatures as large as  $10^6$  K.

A database of over one hundred million EUV bright points has been assembled [16] from the approximately 360,000 EIT full-disk images over the last 12 years.

Marsh et al. (2003) [13] detected conversion from slow magnetoacoustic waves to kink waves, as a 5 minute oscillations were propagating from the chromosphere and transition region into coronal loops, being quickly damped.

### 2.3 Coronal Mass Ejections

An automated method to CME identification able to capture more than 75% of CMEs has been developed by Robbrecht and Berghmans [14].

SOHO UVCS, LASCO, and EIT have obtained a great amount of pictures and movies from CMEs (Figure 3), velocities have been measured. Very usefull CME catalogs have been provided to the scientific community, that permit an easier and better analysis and understanding of solar wind and solar energetic particle events. Studies about CME acceleration, coronal jets, current sheets, etc., have also been performed.

### 2.4 Solar Wind

SUMER observations are described in [15]. They account for solar wind acceleration in funnel-like flux tubes originated in coronal holes and establish the acceleration region between 5 and 20 Mm above solar surface.

McIntosh and Leamon (2005) [16], found solar wind velocities and composition from ACE in good agreement with the photospheric magnetic field measurements from MDI.

Vasquez et al. (2003) [17] and Cranmer and Van Ballegooijen (2005) [18] have studied the differences between fast and slow wind at solar minimum as an effect of the different rates of superradial flux-tube expansion over the polar coronal holes and at the edge of the streamer belt. A combination of the two effects is able to produce the observed differences: a raising of the height of the Parker critical point for flux tubes within about 20 degrees of the closed-field part of the streamer tends to produce a slow wind solution; and the near-streamer flux tubes undergo a lower rate of Alfvén wave dissipation at heights at and above the critical point, which leads to a lower wind speed and higher mass loss rate.

Observationally, there is growing evidence that coronal holes and streamers share at least some of the same microphysical processes.

UVCS observations have driven new interest in collisionless wave-particle resonances, specifically the ion cyclotron resonance, as potentially important mechanisms

for damping wave energy and preferentially energizing positive ions in the accelerating solar wind. But no observations are still made of ion cyclotron waves. There are several theories about the acceleration of ions in coronal holes but the topic remains without observational features that clearly determine which kind of instability may be working in the solar wind acceleration in coronal holes.

Related to solar wind composition the CELIAS MTOF group (at the University of Maryland) has recently identified a new type of event. In these Heavy Ion Depletion Events (HIDEs), He and all the heavier ions are depressed relative to solar wind protons by one to two orders of magnitude. Six of such events (with durations ranging from 6 to 48 hours) have been identified. At least three of these HIDEs occur in the declining phase of high speed streams.

## 2.5 Solar Energetic Particle

Energetic particles during quiet time periods detected by EPHIN have been analysed in [20]. Del Peral et al. [21] analysed the electron spectrum in the energy range 150 keV to 10 MeV, measured by EPHIN sensor on board SOHO observatory during 1996 quiet time periods. The results show that the dominant electron population is of jovian origin with spectral indexes in the range from 1.5 to 1.8. Estimation of the emission intensity of electrons from the jovian magnetosphere is also obtained. Unexpected recurrence of jovian electrons at the middle of 1996 during poor Earth-Jupiter magnetic connection has been observed (Figure 5).

Temporal profiles of energetic ions and electrons observed at 1 AU during solar energetic particle events are mainly determined by particle injection features, the observer location relative to the source region at the Sun, and the interplanetary space plasma and field conditions during particle transport. Gómez-Herrero et al. [19] analysis of the temporal profiles of 18 solar energetic particle events detected by EPHIN have been analyzed by fitting a pulse function to them in order to find a set of parameters which can be used to characterize the events. Composition and energy spectra of the different particle population are obtained for both impulsive and gradual SEP events [22].

The ionic charge distributions of solar energetic particles (SEP) events are an important diagnostic of the plasma and acceleration mechanisms and conditions at the source region in the solar corona. The large difference between gradual,  $Q(\text{Fe}) \sim 14$ , and impulsive,  $Q(\text{Fe}) \sim 20$ , SEP events has not yet been explained. New measurements of ionic charge states with instruments of improved sensitivity over a wide energy range on several spacecraft (SAMPEX, SOHO, ACE) have, however, shown that this picture was oversimplified. In gradual IP-shock related events, the ionic charge state of heavy ions was observed to generally increase with energy, with a large event to event variability. For impulsive events, previous work indicated a systematic but modest increase of  $Q(\text{Fe})$  with energy in the range 180-550 keV/nuc.

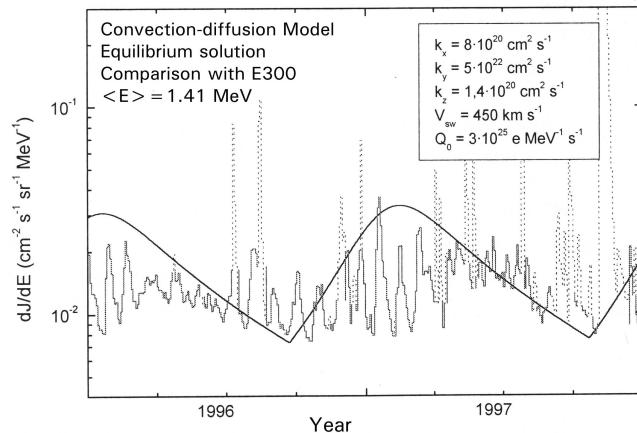


Figure 4: Differential electron flux values plotted in comparison with propagation model predictions (solid line). An unexpected jovian electron super-flux is observed beyond DOY 200 of 1996 during bad magnetic connection between Jupiter and Earth. Dotted lines correspond to SEP events.

However, measurements with SOHO CELIAS STOF now show that the ionic charge of Fe at energies  $\lesssim 100 \text{ keV/nuc}$  in these events is significantly smaller ( $\sim 12.5 \pm 0.9$ ). The large increase of the mean ionic charge of iron in the energy range  $\sim 10 - 550 \text{ keV/nuc}$  as observed with SOHO and ACE can be explained by the establishment of charge interchange during acceleration, as it has been stated by the ESCAPE model Rodríguez-Frías et al. [23], [24], [25] in the dense plasma environment in the low corona, as can be observed in Figure 6 for two impulsive solar events.

### 3 SOHO Scientific Objectives for 2010

The new ESA-NASA missions STEREO, Solar-B, Solar Dynamics Observatory (SDO), Solar Orbiter, etc., give a large number of new and enhanced opportunities. We can only enumerate briefly the future scientific objectives for SOHO in combination with the newer observatories data:

- Subsurface solar weather.
- Quadrature observations.
- Acceleration of the solar wind.

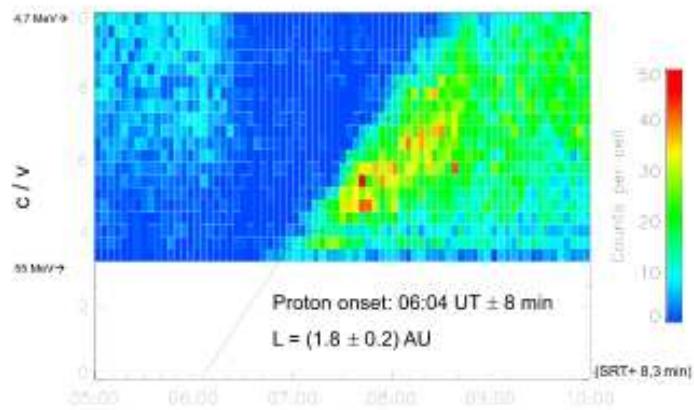


Figure 5:  $\beta-1$  vs time for EPHIN PHA protons during a solar energetic particle event onset.

- CMEs, current sheets, and shocks.
- Dependence of CME propagation on solar cycle phase.
- Reconstruction of CME morphology.
- Multipoint SEP sampling.
- Reconstruction of EUV loop morphology.
- Solar forcing of terrestrial climate.
- Operational information on CME speeds and directions.
- L1 solar wind monitoring under extreme conditions.
- Discovery science.

*Acknowledgements:* This work is supported by the Comunidad de Madrid and Universidad de Alcalá project CCG06-UAH/ESP-0397.

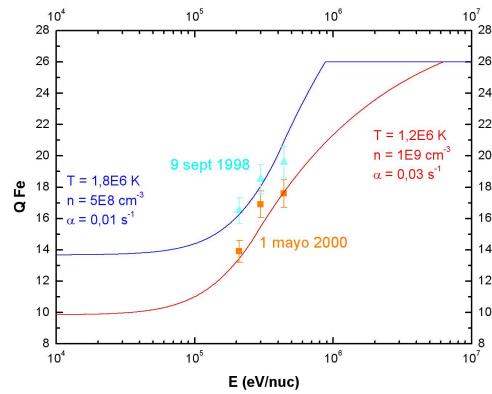


Figure 6: ESCAPE model prediction of the charge state dependence on the energy for two impulsive SEP events.

## References

- [1] Fröhlich, C. 2003, *Metrologia* 40, 60
- [2] Asplund, M., Grevesse, N., and Sauval, A.J. 2005, *APS Conf. Ser.* 226: *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis*, 38, 25
- [3] Lodders, K. 2003, *Meteoritics & Planetary Science*, Vol. 38, Supplement, 38, 5272
- [4] Bahcall, J.N. and Serenelli, A.M. 2005, *ApJ* 626, 530
- [5] Guzik, J.A., Watson, L.S., and Cox, A.N. 2005, *ApJ* 627, 1049
- [6] Couvidat, S., Garca, R. A., Turck-Chize, S., Corbard, T., Henney, C. J., Jimnez-Reyes, S. 2003, *ApJ*. 597, L77
- [7] Dziembowski, W.A. and Goode, P.R. 2005, *ApJ* 625, 548
- [8] Haber, D.A. 2004, in *ESA SP-559: SOHO 14 Helio- and Asteroseismology: Towards a Golden Future*, 676
- [9] Hindman, B.W., Featherstone, N.A., Haber, D.A., Musman, S., and Toomre, J. 2004, in *ESA SP-559: SOHO 14 Helio- and Asteroseismology: Towards a Golden Future*, 460
- [10] Zhao, J. and Kosovichev, A.G. 2004, *ApJ* 603, 776
- [11] Nagata, S. et al. 2003, *ApJ* 590, 1095
- [12] Brosius, J.W. and White, S.M. 2005, *ApJ* 627, 1049
- [13] Marsh, M.S., Wlash, R.W., DeMoortel, I., Ireland, J. 2003, *A&A* 404, L37
- [14] Robbrecht, E. and Berghmans, D. 2004, *A&A* 425, 1097

- [15] Tu, C.-Y., Zhou, C., Marsch, E., Wilhelm, K., Zaho, L., Xia, L.-D., Wang, J.-X. 2005, *Science* 308, 519
- [16] McIntosh, S.W. and Gurman, J.B. 2005, *Solar Phys.* 228, 285
- [17] Vasquez, A.M., et al. 2003, *ApJ* 598, 1361
- [18] Cranmer, S.R., van Ballegoijen, A. A. 2005, *ApJ Suppl.* 156, 265.
- [19] Gómez-Herrero, R. et al. 2005, *Advances in Space Research* 35, 617
- [20] Gómez-Herrero, R. et al. 2000, *Solar Physics* 19, 405
- [21] Del Peral, L. et al. 2003, *Astroparticle Physics* 20, 235
- [22] Gomez-Herrero, R. et al. 2002, *Astroparticle Physics* 17, 1
- [23] Rodríguez-Frías, D., Del Peral, L., Pérez-Peraza, J. 2000, *JGR* 106-A8, 15.657
- [24] Rodríguez-Frías, D., Del Peral, L., Pérez-Peraza, J., 2000, *J. of Physics G: Nuclear and Particle Physics* 26, 259
- [25] Del Peral, L. et al. 2002, *Astroparticle Physics* 17, 415420



# PREPARING ESA Gaia MISSION: STELLAR SPECTRA PARAMETERIZATION BY NEURAL NETWORKS IN THE NEAR INFRARED

DIEGO ORDÓÑEZ BLANCO<sup>1</sup>, JOSÉ CARLOS DAFONTE VÁZQUEZ<sup>1</sup>,  
BERNARDINO ARCAJ VARELA<sup>1</sup>

*1 Departamento de Tecnologías de la Información y de las Comunicaciones,  
Universidade da Coruña, E-15071, A Coruña, SPAIN*

MINIA MANTEIGA<sup>2</sup>

*2 Departamento de CC. de la Navegación y de la Tierra, Universidade de A  
Coruña, E-15011 A Coruña, SPAIN*

ANA ULLA MIGUEL<sup>3</sup>

*3 Departamento de Física Aplicada, Área de Astronomía y Astrofísica, Universidade  
de Vigo, E-36310, Vigo, SPAIN*

**Abstract:** The Gaia mission of the European Space Agency, foreseen to be operative at the beginning of 2012, will extend the Hipparcos legacy by carrying out a census of the Galaxy and providing accurate information about the composition and motion of its main components. Data handling and analysis of information regarding the complete sky down to magnitude 17-18 will be, with no doubt, a challenge for both Astrophysics and Computational Sciences. We present here our preliminary results on the on-going study about the automated derivation of stellar atmospheric parameters in the spectral region of the Gaia RVS (Radial velocity spectrometer) instrument. The use of artificial neural networks (ANN) trained with synthetic model spectra was the method selected for such automated derivation. Both direct stellar fluxes and Fourier transform moduli of them have been considered as inputs to train and test the ANN performance. It is shown that ANN represent a good approach to analyze and parameterize such a large dataset as the one expected from Gaia. Preliminary results achieved are comparable to those obtained by direct spectroscopic or spectrophotometric analysis with synthetic model atmospheres, being their accuracy highly dependent on the spectral signal to noise ratio.

**Keywords:** Stellar spectroscopy – Stellar parameterization – Gaia mission – Artificial Neural Networks.

## 1 Gaia RVS instrument

The Gaia satellite, which is foreseen to be launched near the end of 2011, is one of the present key scientific missions of the European Space Agency (ESA). Gaia mission

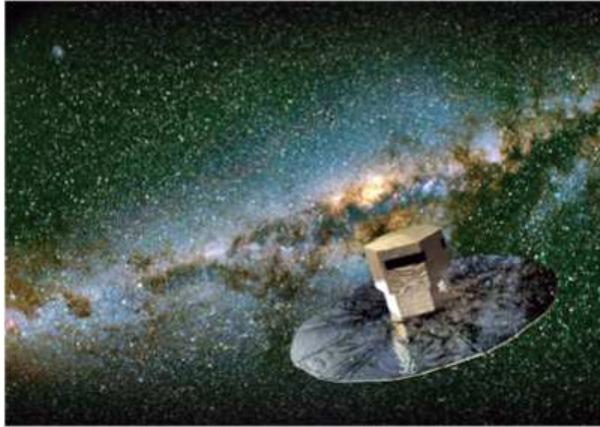


Figure 1: Artist's impression depicting the Gaia spacecraft against a background suggestive of the Milky Way. Figure courtesy of ESA. Illustration by Medialab.

will carry out the most accurate study of the Galaxy components, by compiling exact information on their nature and motion. In the course of its 5 operative years, it will perform precise astrometry (up to 10 microarcsecs at  $V=15$  mag.) and measure the motions of all the sources in the sky plane. This will allow to calculate distances via parallaxes with a precision of 1% for a total of 2.5 million stars down to 2.5 kpc. An artistic view of Gaia spacecraft orbiting the Milky Way can be seen in Figure 1.

Gaia will be equipped with a wide-band spectrophotometer and a radial velocity spectrograph, RVS, that will contribute to the study of the sources' nature and will allow to determine radial motions with precisions between 1-10 km/s for  $V=16-17$  mag. [1]. The RVS domain is the Ca II infrared region, 847-874 nm, a region which is rich in diagnostic lines for the determination of stellar atmospheric parameters, in particular, effective temperatures, surface gravities, overall metallicities and non canonical alpha-elements abundances. This wavelength range has been selected to coincide with the energy-distribution peaks of G and K-type stars which are the most abundant RVS targets. For these late type stars the wavelength interval displays three strong ionized Calcium lines and numerous weak lines, mainly due to Fe, Si and Mg. In early type stars, RVS spectra will be dominated by Hydrogen Paschen lines and may contain weak lines due to carriers as CaII, HeI, HeII and NI. The instrument

operates in time-delayed integration mode, observing each source about 40 times during the 5 years of the mission. Figure 2 shows Gaia RVS spectral domain for a F type star and two values of the signal to noise ratio. Comprehensive information about this project can be found in the Gaia web area at <http://www.rssd.esa.int/gaia>

Over the 5 years mission, RVS will observe around 5 billion transit spectra of the brightest 100-150 million stars on the sky. The on-ground analysis of these spectroscopic data set will be a complex and challenging task, not only because of the volume but because the interdependence of different instruments and modes of observation. As a consequence, data extraction and parameterization should be performed completely in an automatic fashion. These are the reasons why the use of Artificial Intelligence techniques and, in particular, artificial neural networks (ANN), is a good approach to be tested for the case of Gaia-RVS dataset.

## 2 Spectralib: a library of synthetic spectra for Gaia RVS.

Our initial approach consisted of performing simulations on stellar parameter extraction by means of synthetic spectra. For these first tests it was used the Gaia RVS *Spectralib*, a library of 9285 stellar spectra compiled by A. Recio-Blanco and P. de Laverny from Niza Observatory, and B. Plez from Montpellier University. The spectra are based on the new generation of MARC models from The Uppsala Observatory in the RVS region. A technical note describing the models used for the atmospheres from which the synthetic spectra were calculated and what parameters were used is available ([2]). This set of spectra will be named RVS1. More recently, tests were performed with an updated new set of 7168 spectra based in the same library of MARC models specifically computed for RVS Gaia simulations (from now on RVS2 set of spectra). Both grids are essentially equivalent and consist on spectra corresponding to effective temperatures between 4000 and 8000 K (step 250K),  $\log g$  between -1.0 to 5.0 (step 0.5 dex, range -0.5 to 5.0 dex for RVS2), and overall metallicities between -5.0 and 1.0 (with variable step from 1.0 to 0.25 dex). For each model atmosphere, alpha-elements abundance variations of at least the following values: +0.4, +0.2, +0.0, -0.2 dex, were considered with respect to the original abundances in the models. The grid closely covers HR diagram positions for A5, F, G and K stars. Access to RVS-Spectralib is open via the ESA Gaia web pages (<http://gaia.esa.int/spectralib/>).

## 3 The use of ANN for spectral parameterization

Astrophysics is moving towards a more rational use of the costly observational material by means of the intelligent exploitation of large terrestrial and spatial astro-

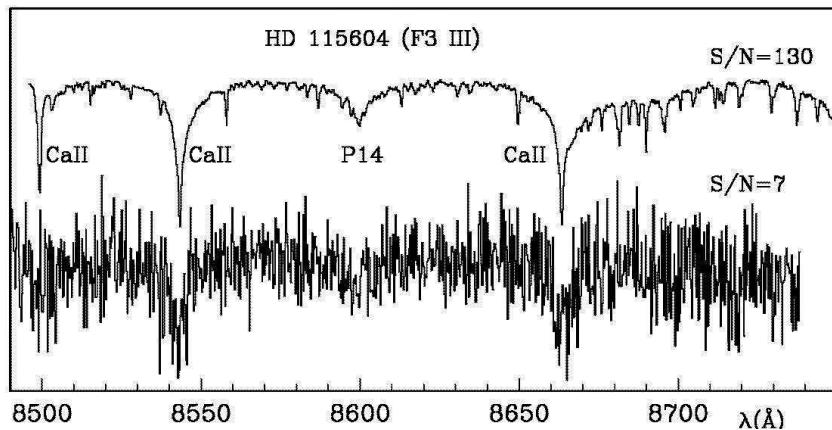


Figure 2: RVS spectral region showing the principal features for a F3III star and two SN levels.

nomical databases. Nowadays, every single project on astronomic instrumentation includes the creation of data archives and their future exploitation, with automatic or pseudo-automatic analysis tools. Several international projects are currently working on management systems for astronomical information by means of the common exploitation of all the available spatial archives and astronomical databases (international virtual observatories initiatives). Our research group is a member of the Spanish Virtual Observatory ([3]), which consists of various networked Spanish groups and aims, among other things, at developing applications for the automatic analysis of astronomical data. AI techniques are among the techniques that are being used for those developments.

Among the different techniques of Artificial Intelligence, ANN have already proved their success in classification problems: they are generally capable of learning the intrinsic relations that reside in the patterns with which they were trained. Some well-known previous works have applied this technique to the problem of stellar spectral parameterization, obtaining different grades of resolution in the extraction of parameters  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$ . Most authors have used multilayer neural networks (MLP) with supervised learning algorithms based on the backpropagation of the training error. As examples we can mention the already classical works by [4], [5], [6], and [7], as well as the more recent works by [8] and [9]. A summary of the current status of automated stellar classification techniques and achievable accuracies can be found in the reviews by [10] and [11]. The range of accuracies ( $1\sigma$ ) obtained in the derivation of the atmospheric parameters are of the order of 150-200 K, 0.3-0.4

dex and 0.15-0.2 dex for  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$ , respectively.

In order to probe the adequacy of RVS-spectral region for the derivation of stellar atmospheric parameters, we choose to train ANN with the ad-hoc calculated synthetic spectra already introduced in the previous section. Tests with ANN were performed in two stages. In an initial phase we wanted to study how signal to noise ratio influences the ANN training phase and parameter determination errors for the case of  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$  extraction. In this phase the RVS1 set of spectra was considered for the tests. In the second stage, we were aimed to consider the performance of the ANN when a different transformed domain for the input spectra is taken into account. Instead of feeding the ANN with the continuum subtracted spectra, their Fourier transformed absolute values (showing the energy distribution of frequencies among the spectra) were considered as the input information to train the networks and validate the parameters values. In this second phase the RVS2 set of spectra was considered for the tests.

### 3.1 RVS spectral parameterization. Signal to noise dependence.

A simple network architecture was chosen for the tests: a feedforward ANN with 1004 input nodes (the number of pixels in the spectra), 1 hidden layer with 150 nodes, and 4 output layers providing the atmospheric parameters,  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$ .

A total of 1764 spectra well distributed in the space of parameters were considered in the training set, while tests were performed on subsets of 465 spectra. Typically, a good performance in the network convergence and low parameter errors were achieved after about 10000 training cycles, which translates to about 24 hours of computational time on an AMD 64 computer (AMDx64 Dual Core 3200GHz 2GB RAM).

Gaia-RVS spectra will be of very different quality depending mostly on the stellar brightness. It has been proposed that the end-of-mission SN ratio for a typical star in the Galaxy, a G5V with  $V=15.5$  or a F2II with  $V=14.5$  will be about 10. In order to have into account the effect of the SN values in the ANN performance, we have delivered tests taking into account six values of the SN ratio: 10, 25, 50, 75, 100 and 500. The model of noise considered was a simple gaussian white noise.

As expected, it was found that the mean errors in the extraction of the parameters depend on the SN ratio. It was also found that better results were achieved when the training set contains spectra with a similar distribution of noise as the validation set. Validation on spectra displaying different SN provides different degrees of accuracy for the atmospheric parameters. We chose to train the networks with a training set composed by spectra showing a distribution of SN ratios among 10 to infinite, representative of the whole sample to be validated. Mean errors as low as 50.0 K, 0.10, 0.09 and 0.06 dex for effective temperatures,  $\log g$ , metal abundance and  $[\alpha/\text{Fe}]$ , respectively, were found for ANN tested on synthetic spectra with no noise added.

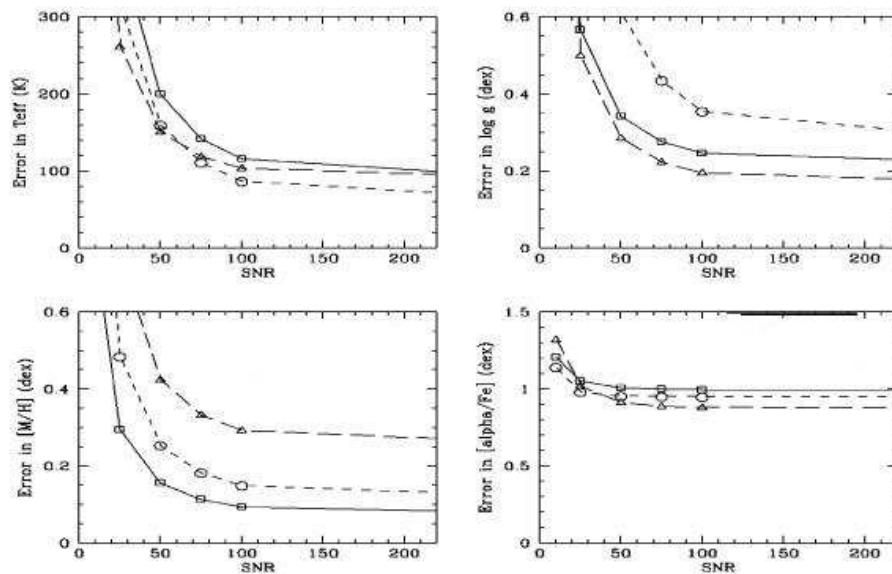


Figure 3: Results on the parameters extraction as a function of S/N, and for different stellar populations as follows: squares (cool dwarfs, metal rich), circles (cool giants, intermediate metallicity) and triangles (hot subgiants, very metal poor).

The errors grew to values of 74 K, 0.17, 0.13 and 0.08 dex in the case of SN 100; 115 K, 0.24, 0.19 and 0.11 dex for SN 50; 191 K, 0.37, 0.30 and 0.16 dex for SN 25 and finally to 391 K, 0.77, 0.55 and 0.25 dex for SN 10.

Figure 3 shows the performance of the ANN as a function of SN when considering validation on 3 set of spectra representative as tracers of three galactic stellar populations: cool dwarfs, metal rich stars ( $T_{\text{eff}} 5000\text{-}6000\text{K}$ ,  $\log g \geq 3.5$  dex and  $[\text{Fe}/\text{H}] \geq -0.5$  dex); cool giants, intermediate metallicity ( $T_{\text{eff}} \geq 5000$  K,  $\log g = 1.0\text{-}3.5$  dex,  $[\text{Fe}/\text{H}] \geq -1.0$  dex); and hot subgiants, very metal poor stars ( $T_{\text{eff}} \geq 6000\text{K}$ ,  $\log g = 2.0\text{-}4.0$  dex,  $[\text{Fe}/\text{H}] = -1.0\text{-}2.5$  dex). From the data in the figures, it is obvious that the SN heavily influence the learning process. In our experience, essentially poor results are encounter when training and testing spectral samples with different SN values.

### 3.2 RVS spectral parameterization in the Fourier transformed domain.

RVS2 set of spectra needed to be preprocessed before transformed to the Fourier domain. The procedure consisted in the following steps: continuum fit by linear regression and continuum subtraction, subtraction of the mean value and multiplication by a hamming window. Then, calculation the Fast Fourier Transform (FFT) algorithm was applied and the absolute values of the FFT determined. As in the previous experiment, one network was trained for each parameter ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$ ). Again, feed-forward ANN fully connected were considered and for the training phase a standard on line backpropagation algorithm. The dimension of the input layer was the same as the number of points in the processed spectra (400 points after preprocessing and filtering), and the output layer had a unique process element to obtain each of the parameters.

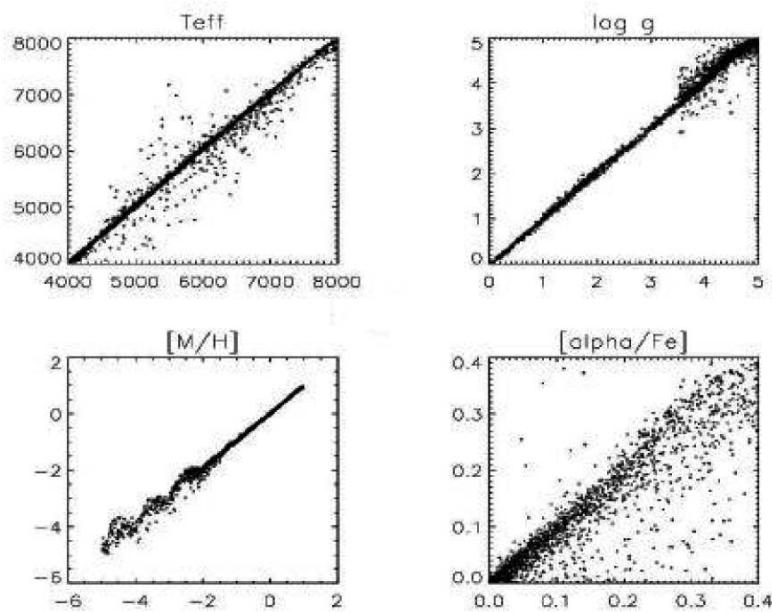


Figure 4: Astrophysical parameters recovering with ANN trained in the Fourier domain.

The method was found to be very robust. Astrophysical parameters can be determined with the following mean errors :  $T_{\text{eff}}$  54 K,  $\log g$  0.08 dex,  $[\text{Fe}/\text{H}]$  0.06 dex

and  $[\alpha/\text{Fe}]$  0.03. Ruling out the 20% worse cases, errors in the following range were found:  $T_{\text{eff}}$  -35 to 62 K,  $\log g$  -0.05 to 0.16 dex,  $[\text{Fe}/\text{H}]$  -0.01 to 0.12 dex, and  $[\alpha/\text{Fe}]$  -0.07 to 0.1 dex. When compared with the previous experiment (RVS1 set with no noise added) it can be seen that an improvement in the parameterization of  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $[\alpha/\text{Fe}]$  has been achieved, while essentially the same errors are found in the recovery of  $T_{\text{eff}}$ .

The performance of the method is shown in Figure 4 where the error distribution for the recovery of each four parameters is shown, while in Figure 5 it is shown a gaussian fit for the central, best behaved population.

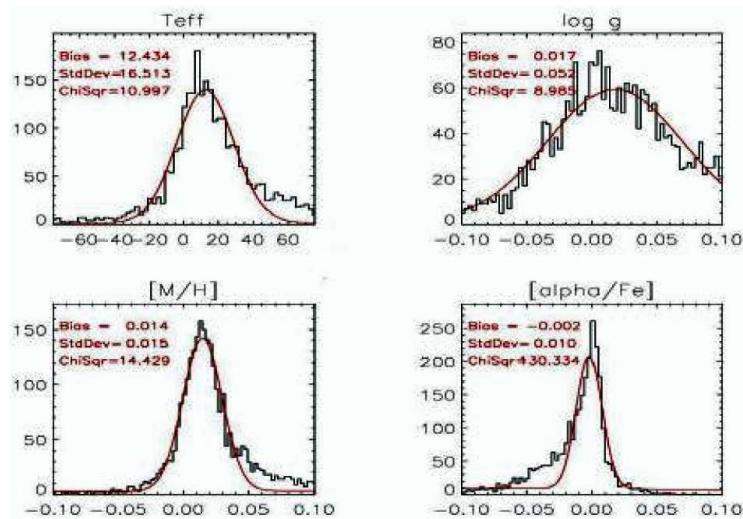


Figure 5: Distribution of errors for derivation of astrophysical parameters with ANN trained in the Fourier domain.

## 4 Conclusions and future work

We presented our first results on the automatic derivation of stellar parameters in the RVS spectral region, by the use of artificial neural networks trained with synthetic model spectra. The results achieved are comparable to those obtained by the use of spectrophotometry, being the accuracy highly dependent on the signal to noise ratio of the spectra. The training of ANN in the FFT domain provides a novelty factor, and the good results found in the performed experiment open a new way to be explored in the future with noise-added spectra.

Our results show that ANN can be a good approach to extract atmospheric parameters from Gaia-RVS spectra, providing that the SN ratio of the training and testing spectral set be well characterized. Mean errors as low as 115 K, 0.24, 0.19 and 0.11 dex for effective temperatures, logg, metal abundance, and  $[\alpha/\text{Fe}]$ , respectively, were reached for synthetic spectra with SN 50.

*Acknowledgements:* This work is supported by the Spanish MEC ESP2006-13855-C02-02 and FEDER funds.

## References

- [1] Katz, D., Munari, U., Cropper, M. et al. 2004, MNRAS 354, 1223
- [2] Recio Blanco, A., de Laverny, P., Plez, B. 2005, European Space Agency technical note RVS-ARB-001
- [3] Solano, E. 2006, Lecture Notes and Essays in Astrophysics, vol. 2, 71
- [4] von Hippel, T., Storrie-Lombardi, L.J. et al. 1994, MNRAS 269, 97
- [5] Singh, H.P., Gulati, R.K. and Gupta, R. 1998, MNRAS 295, 312
- [6] Bailer-Jones, C.A.L., Irwin, M., Gilmore, G. et al. 1997, MNRAS 292, 157
- [7] Vieira, E.F., Ponz, J.D. 1995, A&AS 111, 393
- [8] Gupta, R., Singh, H.P., Volk, K. et al. 2004, ApJS 152, 201
- [9] Willemsen, P.G., Hilker, M., Kayser, A. et al. 2005, A&A 436, 379
- [10] Bailer-Jones, C.A.L. 2002, in “Automated Data Analysis in Astronomy”, R. Gupta, H.P. Singh, C.A.L. Bailer-Jones (eds.), Narosa Publishing House, New Delhi, India. Preprint: arXiv:astro-ph/0102223v1
- [11] Allende Prieto, C. 2004, Astronomische Nachrichten, Vol.325, Issue 6, 604

## Author Index

- Aguado, J., 163  
Angelats i Coll, M., 151  
Arcay Varela, B., 225
- Battaner, E., 83  
Beckman, J. E., 177  
Bernabéu, G., 189  
Bravo, E., 45
- Canto, J., 201  
Castro-Tirado, A. J., 131  
Cerrato, Y., 163  
Cid, C., 163  
Collados, M., 113  
Cruz, R., 213
- Dafonte Vázquez, J.C., 225  
Del Peral, L., 213  
Domínguez, I., 45  
Domingo, M., 201  
Drossart, P., 13
- Florido, E., 83  
Forget, F., 151
- Gómez-Herrero, R., 213  
Gomis, O., 201  
González-Galindo, F., 151  
Guijarro, A., 83  
Gutiérrez, L., 177
- Höflich, E., 45  
Heber, B., 213  
Hebrero, G., 213  
Hueso, R., 1
- Jordi, C., 33
- López-Valverde, M. A., 13  
López-Valverde, M.A., 151  
Luna, R., 201
- Müller-Mellin, R., 213  
Má rquez, I., 63  
Manteiga, M., 225  
Masana, E., 33  
Millan, C., 201
- Ordóñez Blanco, D., 225
- Pérez-Hoyos, S., 1  
Peralta, J., 1  
Piccioni, G., 13  
Piersanti, L., 45
- Rodes-Roca, J. J., 189  
Rodríguez-Frías, D., 213  
Rubiño-Martín, J.A., 83  
Ruiz-Granados, B., 83
- Sánchez, F., 103  
Sánchez-Lavega, A., 1  
Saiz, E., 163  
Santonja, M. C., 201  
Satorre, M. A., 201  
Straniero, O., 45
- Tornambe, A., 45  
Torrejón, J. M., 189
- Ulla, A., 225
- Vilaplana, R., 201
- Wimmer-Schweingruber, R., 213
- Zurita, A., 83



## Subject Index

- ΛCDM, 89
- AGN, 84
- alpha elements abundance, 227
- antitruncation, 92, 93
- artificial neural networks, 227
  
- bars, 84
- bolometer, 201
- break, 91, 94, 96
- bubbles, 84
  
- clusters, 86
- CMB, 86, 98
- Coma cluster, 86
- comets, 203
- compact object, neutron star, 190
- Cosmological constant, 46
- Cosmological Principle, 86
- CRSF, 195
- CRSF, optical depth, 194, 195
- CRSF, variability, 194
- CRSF, width, 194
- cyclotron line, 194
  
- Dark Energy, 58
- Dark energy, 46, 48
- dark energy, 45, 48
- dark matter, 87, 88
- DDO 154, 90
- diffusion, 97
- Dst index, 163, 170–173
- Dust storms, 152
- dwarf galaxies, 90
- dwarf spirals, 87
- dynamo, 84
  
- elliptical galaxies, 89
- escape, 88, 93
  
- Faraday rotation, 86
- filaments, 86
- filling factor, 182
- flaring, 88, 98
- Fourier transform, 231
  
- Gaia RVS spectrograph, 226
- Gaia spacecraft, 226
- galaxies, back holes, 65
- galaxies, bars, 63–65, 69–72, 74, 76, 78
- galaxies, HII, 73
- galaxies, interactions, 63–65, 68, 73, 78, 79
- galaxies, LINER, 73
- galaxies, nuclear activity, 63, 69, 73
- galaxies, nuclear bars, 72, 78, 79
- galaxies, nuclear disks, 72, 76
- galaxies, nuclear rings, 78
- galaxies, nuclear source, 72
- galaxies, nuclear sources, 65
- galaxies, nuclear spirals, 72
- galaxies, Seyfert, 65, 72, 73
- galaxies, starbursts, 71
- Galaxy astrometry, 226
- gamma-ray bursts, 143
- General Circulation Models, 151, 153–155, 157, 159
- General Circulation Models, Thermosphere, 154
- geomagnetic storm, 164, 166–168, 170–172
- Globular clusters, 46
  
- halo, 96
- halo, DM, 96
- hazards, 168, 171
- helicity, 84
- HI, 84

HII regions, 84, 177  
 HII regions, inhomogeneous models, 182, 185  
 HII regions, ionization process, 178  
 HII regions, photon escape, 178  
 HII regions, ionization process, 179  
 Hubble Deep Field, 98  
 Hubble diagrams, 47, 58  
 IC10, 90  
 ice density, 208  
 ices, interstellar medium, solar system, 201  
 infrared spectroscopy, 201, 208  
 intracluster medium, 86  
 ion irradiation, 204  
 jets, 84  
 laboratory experiments, 201  
 large scale structure, 86  
 lenticular, 96  
 Local Group, 90, 97  
 lopsidedness, 83, 98  
 M31, 85, 97  
 M33, 97  
 M82, 85  
 M83, 85  
 magnetic arms, 84  
 magnetic diffusion, 86, 90, 97  
 magnetic field, CRSF, 189  
 magnetic field, cyclotron resonant scattering feature, 189  
 magnetic fields, cyclotron resonance scattering features, 192  
 magnetic stars, 85  
 magnetic tension, 88, 89  
 Mars atmosphere, 151, 152, 154, 155  
 Mars atmosphere, General Circulation Models, 155  
 Mars atmosphere, Thermosphere, 154, 156, 158, 159  
 Milky Way, 89, 97  
 molecular cloud, 94  
 molecular clouds, 84  
 MOND, 88  
 NGC2654, 92  
 NGC4449, 90  
 NGC5055, 94  
 NGC6504, 91  
 NGC6946, 85, 89  
 outflow, 90  
 persistent X-ray emission, 190  
 planck, 98  
 polarization, 85  
 pregalactic medium, 84  
 pulse phase spectroscopy, 194  
 Quantum vacuum energy, 46  
 quasars, 86  
 radio lobes, 84  
 reconnection, 84  
 reversals, 84  
 ring, 89, 90  
 robotic astronomical observatories, 131  
 robotic astronomical observatory, 131  
 rotation curve, 83, 87–89, 94  
 rotation curves, 87  
 solar chromosphere, heating, 123  
 solar instruments, IBIS, 125  
 solar instruments, POLIS, 125  
 solar instruments, TESOS, 125  
 solar instruments, TIP, 125  
 solar instruments, VIP, 125  
 solar telescopes, 114  
 solar telescopes, adaptive optics, 119  
 solar telescopes, ATST, 125  
 solar telescopes, Dutch Open Telescope, 118, 124  
 solar telescopes, EST, 120, 126

solar telescopes, EST science cases, 120  
solar telescopes, Hinode, 126  
solar telescopes, Sunrise, 126  
solar telescopes, Swedish Solar Telescope, 118, 119, 124  
solar telescopes, THÉMIS, 115, 125  
solar telescopes, Vacuum Tower Telescope, 115, 119, 125  
Space weather, 163, 165, 168–170, 172, 173  
spiral arms, 84  
spiral galaxies, 85, 87, 90, 91, 96, 98  
star, neutron star, 190  
starburst galaxies, 85  
stars, black hole candidate, 189  
Stars, evolutionary models, 53, 54, 58  
stars, neutron star, 189  
stars, pulsations, 216  
Stars, rotating WDs, 56  
Stars, WDs, 53, 54  
stars, white dwarf, 190  
stellar parameterization, 227  
Sun, 85  
sun, active regions, 120, 123  
sun, Alfvén speed, 122  
sun, basic science, 114  
sun, cut-off frequency, 122  
sun, EAST, 120  
sun, ephemeral regions, 121  
sun, magnetic fields, 119  
sun, magnetic flux removal, 121  
sun, magneto-acoustic waves, 122  
sun, moving magnetic features, 121  
sun, near-infrared, 115  
sun, network, 121  
sun, numerical magnetohydrodynamical simulations, 116  
sun, planetary and astrophysical connection, 114  
sun, polarisation, 119  
sun, quiet sun, 121  
sun, radiative transfer, 117  
sun, reconnection, 121  
sun, solar-terrestrial connection, 113  
sun, spectropolarimetry, 115  
sun, sunspot decay, 121  
sun, wave propagation, 122  
supernovae, 50  
Supernovae, calibration, 47  
Supernovae, dependence on host galaxy, 49  
Supernovae, evolution, 49  
Supernovae, explosion models, 58  
Supernovae, extinction, 49  
Supernovae, extinction corrections, 55  
Supernovae, K-corrections, 48  
Supernovae, peculiar, 48  
Supernovae, rotation, 58  
supernovae, SNIa, 45  
Supernovae, thermonuclear explosions, 51  
telescope control operating system, observatory managers, 139  
temperature desorption, 208  
Thermosphere, 152, 153, 155  
thermosphere, 151  
Thermosphere, General Circulation Models, 156  
threshold, 94  
truncation, 83, 91–94, 96  
Tully-Fisher relation, 90, 93  
UV irradiation, 204  
Venus, Venus atmosphere, Venus Express, VIRTIS, 1  
Virial theorem, 88  
Virtual Observatories, 228  
walls, 86  
warp, 95, 96  
WMAP, 98

worldwide robotic astronomical telescope  
network, 147

X-ray binaries, Be/X-ray binary, 190

X-ray binaries, Cataclysmic Variables,  
190

X-ray binaries, high mass X-ray binary,  
189

X-ray binaries, HMXB, 190

X-ray binaries, IMXB, 190

X-ray binaries, LMXB, 190

X-ray binaries, SFXT/X-ray binary, 191

X-ray binaries, SG/X-ray binary, 190

X-ray binaries, X-ray pulsars, 191

X-ray observatory, Ginga, 193

X-ray observatory, INTEGRAL, 191

X-ray observatory, RXTE, 189, 192, 193

