Solar Physics and Solar Eclipses

Proceedings of an International Symposium held at
Waw an Namos, Libya, 27-29 March 2006

Edited by

RENZO RAMELLI

Istituto Ricerche Solari Locarno
Locarno, Switzerland

OSAMA M. SHALABIEA

Department of Physics, Faculty of Science, Sebha University
Sebha, Libya

and

Faculty of Science, Cairo University
Cairo, Egypt

IBRAHIM SALEH

Faculty of Engineering, Al-Fateh University
Tripoli, Libya

JAN O. STENFLO

Institute of Astronomy, ETH Zurich
Zurich, Switzerland

CONTENTS

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

INTRODUCTORY REMARKS BY THE CHAIRMAN OF THE SCIENTIFIC ORGANIZING COMMITTEE

The International Symposium on Solar Physics and Solar Eclipses (SPSE 2006) is unique in several respects:

- It takes place in a remote location of the Sahara desert, where the participants will have the world's best viewing conditions of the 29 March 2006 solar eclipse.
- It is a model for collaboration between scientists from Arab and western countries.
- It brings together, for the first time, scientists representing the two disciplines of solar astronomy and solar energy utilization.
- It is the first major international conference in Libya in this field of science.

The organization of the symposium has been a difficult enterprise, due to the many novel aspects that we had never encountered before. The idea to do something like this came from Dr. Osama Shalabiea of Sebha University, who contacted me last summer about it. At that time I was quite skeptical that such a conference at a remote location in the desert without any preexisting infrastructure could be successful, but the unique circumstances made it irresistible to try. I am glad we did. Although at times it seemed that the whole enterprise was at the brink of collapse, last-minute solutions to the various problems could always be found thanks to the dedication and determination of our colleagues at Sebha University. This has now resulted in a rich, high-quality symposium program that is exceeding our expectations.

This enterprise would not have been possible without the full, determined backing that we at all times have received from Sebha University, including the university President, the Local Organizing Committee, and in particular thanks to the enormous efforts invested by the coordinating scientist and astronomer at Sebha, Osama Shalabiea. I also want to thank the members of SOC and all the symposium participants for their patience and faith in the success of the meeting during times when the situation was still ill defined. It is because all of you have decided to come together at Waw an Namos to contribute to the meeting with high-quality presentations that our enterprise is becoming a great success. Finally we are grateful to Rob Schreuders of Eclipse-City and all his coworkers for building the infrastructure and finding practical solutions to all our individual problems, while working under difficult conditions and time pressure.

Personally I see this meeting not as an end product, but as a new beginning, a new chapter of international scientific collaboration that will be followed up in various forms after we have departed from Waw an Namos. Let us all work together for this to happen!

Jan Olof Stenflo


الندوة الدولية حول فيزياء الشمس والكسوف الشمسي
International Symposium on Solar Physics and Solar Eclipses
تنظمها جامعة سبها بالتعاون مع المركز الليبي للاستشعار عن بعد وعلوم الفضاء
في الفترة من 27- 29- الربيع (مارس) 2006 بمنطقة واو الناموس - ليبيا
27-29 March 2006, Waw An Namos, Libya
Organized by: Sebha University &
The Libyan Center for Remote Sensing and Space Science (LCRSSS)
IHY 2007


تنظمها جامعة سبها بالتعاون مع المركز الليبي للاستشعار عن بعد وعلوم الفضاء
في الفترة من 27- 29- الربيع (مارس) 2006 بمنطقة واو الناموس - ليبيا
29 March 2006, Waw An Namos, L
Organized by: Sebha University
The Li r for Remote Sensing and Space CRSSS)
IHY 2007

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

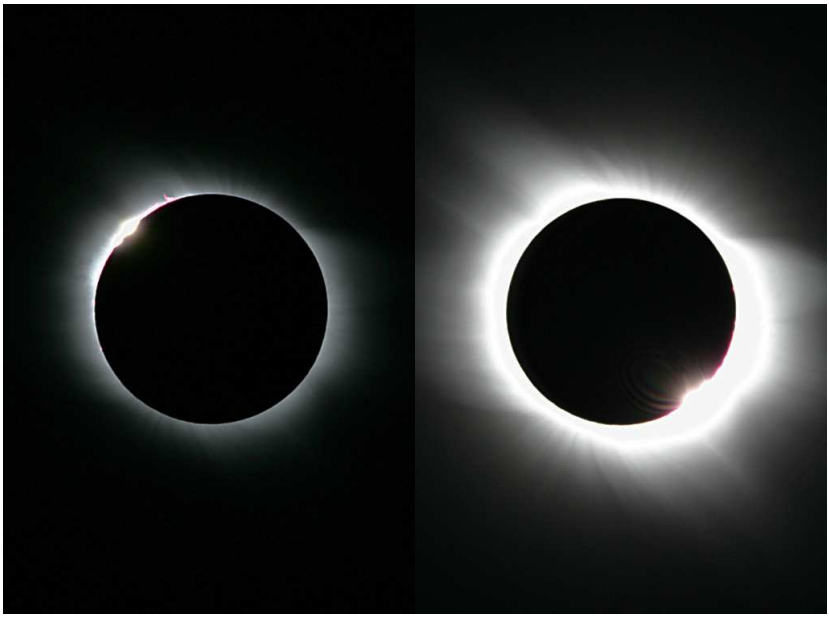
INTRODUCTORY REMARKS BY THE CHAIRMAN OF THE LOCAL ORGANIZING COMMITTEE

From the first moment I was nominated as the chair of the Local Organizing Committee (LOC) of the International Symposium on Solar Physics and Solar Eclipses (SPSE 2006), I realized with my LOC colleagues the great challenges we were going to have, in particular with the time and the place where we planned to organize the symposium: Waw an Namos. That was the region the scientific committee chose as the best location to observe the total solar eclipse event and to gather the scientists from all over the world for the symposium as well. Waw an Namos is in the middle of the Libyan Sahara desert where no infrastructure is available, specially the electric power, water and communication, in addition it is difficult to reach the area by normal transportation means. On the other hand, we were very restricted from the point of view of the time of the eclipse and there was no way to change or even shift this event. At the same time, these challenges gave us the power to work hard to overcome the problems we faced, to ensure the success of the symposium and to achieve our goals. Indeed, this symposium was unique in its subject, time and place. It has been a great honor to organize this meeting in parallel to such a special event as a total solar eclipse; it was indeed the first time that this was done in Libya.

Besides all of our efforts, we received the full support from the president of Sebha University, which we appreciate and acknowledge very much. We thank also the Winzrik Company for the work they did to overcome the difficulties of accommodation in such a remote area.

As a conclusion I see this meeting as a special new experience for all the staff at Sebha University. Moreover, it was a new beginning toward a closer international scientific collaboration in these new fields of solar physics and solar energy sciences. Indeed all measurements, the time and location of the meeting, represented a big challenge, and at the end all was a great success.

Ahmaid O. Zaidan



International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

POLARIZATION AT THE EXTREME LIMB OF THE SUN AND THE ROLE OF ECLIPSE OBSERVATIONS

JAN O. STENFLO
Institute of Astronomy, ETH Zentrum, CH-8092 Zurich, and
Faculty of Mathematics & Science, University of Zurich

Abstract. Scattering of light in the Sun's atmosphere produces polarization throughout the solar spectrum. Due to the changing scattering geometry this polarization increases monotonically from disk center towards the solar limb. Many different factors influence the magnitude of the polarization: The intrinsic polarizability of the scattering atomic system (combination of angular momenta and nuclear spin for the involved atomic levels, quantum interferences and optical pumping), collisional processes and depolarizing magnetic fields, the ratio between line and continuum opacities, anisotropies of the radiation field, small-scale structuring of the atmosphere (deviations from plane-parallel stratification), etc. As all these effects vary greatly from line to line, the linearly polarized spectrum becomes richly structured, in a way that has little resemblance to the structuring of the intensity spectrum. According to a classic solution by Chandrasekhar for an idealized, purely scattering plane-parallel atmosphere the linear polarization reaches asymptotically at the extreme limb the very high value of 11.7 %. Measurements a few arcsec inside the limb show much smaller polarization values, generally much below 1 %, and in most cases less than 0.1 %. The polarization amplitude is however expected to increase dramatically at the extreme limb, within the last arcsec, and could possibly approach the Chandrasekhar limit in some spectral lines. Observations of this increase have not been possible in the past due to seeing effects in the Earth's atmosphere, telescope resolution, and stray light in the instrument. All these limitations can be avoided at a total solar eclipse, when the edge of the Moon is used as a knife edge to isolate the extreme limb from the glaring radiation of the solar disk, to allow us to record the polarization from the separate height layers of the atmosphere, from the photosphere through the chromosphere. This has to be done during the brief flash phase of the eclipse (the transition to totality at second contact, or the emergence from totality at third contact). The polarization of the flash spectrum has never been successfully recorded in the past but is being done by a Swiss team from ETH Zurich during the 29 March 2006 eclipse at Waw an Namos, Libya.

Key words: polarization, coherent scattering, solar eclipse, flash spectrum

1. Introduction

Due to coherent light scattering in the Sun's atmosphere the solar spectrum gets linearly polarized, even in the absence of any magnetic fields. In recent years it has been found that this linearly polarized spectrum is as richly structured as the ordinary intensity spectrum, but that the spectral structures are very different and governed by other physical processes than those of the ordinary intensity spectrum (Stenflo, 2004). Since we are dealing with a very different and previously unfamiliar spectrum, the linearly polarized spectrum is referred to as the "Second Solar Spectrum" (Stenflo & Keller, 1996; Stenflo & Keller, 1997). The amplitude of the scattering polarization is found to increase steeply as we approach the solar limb.

In radiative-transfer theory there is a classic analytical solution by Chandrasekhar (1950) for an ideal, purely scattering, plane-parallel atmosphere, which shows how

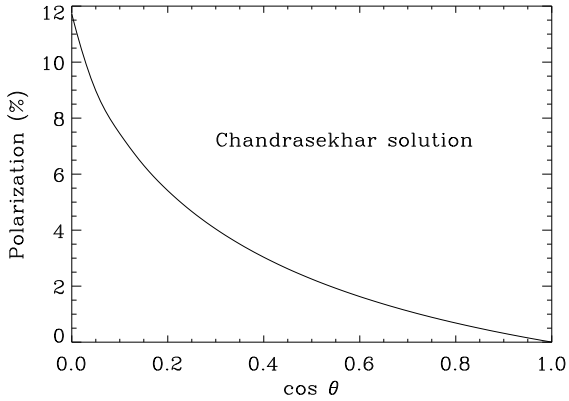


Fig. 1. Degree of linear polarization produced by an idealized plane-parallel atmosphere, in which the radiative transfer is governed exclusively by classical dipole-type scattering (like Thomson scattering at free electrons). The plane of polarization is oriented perpendicular to the radius vector (the line from disk center to the chosen disk position). The center-to-limb distance is given in terms of $\mu = \cos\theta$, where θ is the angle between the vertical direction and the line of sight (the heliocentric angle). The polarization increases from zero at disk center ($\mu = 1$) to 11.7 % at the extreme limb ($\mu = 0$).

the polarization amplitude increases to asymptotically 11.7 % at the very edge of the Sun's limb. Although this idealized treatment gives us general guidance, the real Sun differs in various ways. The real atmosphere is far from purely scattering, since collisional effects (collisional excitation, deexcitation, broadening, and depolarization) and competing, non-polarizing opacity sources play a significant role, and these effects are height dependent. The various Fraunhofer lines are formed at different heights in the atmosphere. The position of the limb depends on wavelength. For photospheric spectral lines the limb lies below the chromosphere. The structure of the chromosphere at the limb is characterized by a forest of spicules (plasma jets that shoot out from below), which means that the plane-parallel concept becomes meaningless. We thus see that the way in which the scattering polarization varies asymptotically near the extreme limb depends on the height and physics of the line formation process, which is individual for each spectral line considered, and it also depends on the microstructuring of the solar chromosphere.

At second contact during a total solar eclipse we have a transition to totality, when the Moon covers up the last sliver of the Sun's disk. As the Moon moves about $\frac{1}{2}$ arcsec per second (360 km/s on the Sun), the flash phase only lasts on the order of 10 s. During this brief time the Sun's spectrum changes dramatically, from the usual absorption-line spectrum to an emission spectrum, since the continuous spectrum vanishes in the chromosphere, and only the strong chromospheric resonance lines stand out in emission, together with the coronal forbidden lines. With an experiment for the 29 March 2006 eclipse at Waw an Namos, Libya, we (Feller *et al.*, 2006) aim to capture the polarization of this emission line spectrum over the whole visible spectral range, from the UV to the near IR, with high time resolution (about 50 frames per second), which corresponds to the phenomenal height resolution on the Sun of about 10 km.

2. Polarization of the continuum

Real stellar atmospheres behave very differently from the idealized Chandrasekhar case, since many different competing opacity sources contribute to the formation of the spectrum, and the relative weights of the various contributions are strongly height dependent. The dominating opacity source for the Sun's continuous spectrum in the visible range is H^-, which acts as pure absorption and does not generate any polarization. All the polarization of the continuum can be attributed to two processes: (1) Rayleigh scattering at neutral hydrogen, and (2) Thomson scattering at free electrons. Both processes behave like classical dipole scattering.

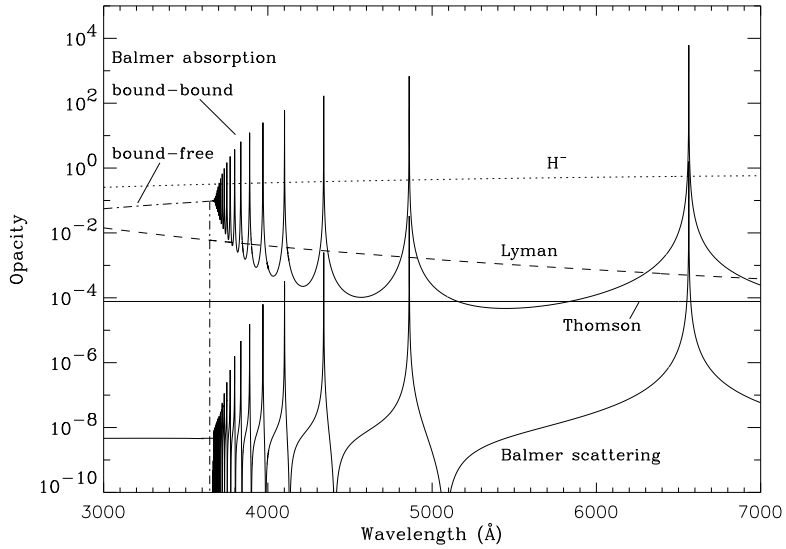


Fig. 2. Opacity contributions (in arbitrary units) to the visible part of the Sun's continuous spectrum, from Stenflo (2005). The diagram is representative of the temperature-density conditions at a height in the solar atmosphere where the Sun's spectrum at a disk position of $\mu = 0.1$ is predominantly formed.

The physics behind the formation of the Sun's visible continuum is almost exclusively determined by the interaction between radiation and hydrogen plus free electrons. The contributions from other chemical elements are insignificant. Figure 2 from Stenflo (2005) gives an overview of the opacity sources between 3000 and 7000 Å. As the relative magnitudes of the various opacity sources depend on temperature and density, this type of diagram is strongly height dependent. Figure 2 is representative for the heights in the solar atmosphere, where the spectrum at a disk position of $\mu = 0.1$ (about 5 arcsec inside the limb) is formed.

The vast majority of the neutral hydrogen atoms are in the ground state. Resonant transitions from the ground state would result in the EUV lines of the Lyman series, but for photons at visible wavelengths we are far from these resonances. Still scattering from the ground state can be viewed as scattering in the Lyman series lines, although in the distant line wings. Since the scattering probability in the dispersion wings drops off fairly slowly, as $\Delta\lambda^{-2}$, with distance from the respective

resonances, this Rayleigh scattering (labeled "Lyman" in Fig. 2) is still very significant at visible wavelengths and the second most important opacity source after H^- absorption.

A competing opacity in the visible is provided by the bound-bound transitions in the Balmer series (that start from the second, excited level in hydrogen, with $n = 2$, which however is orders of magnitude less populated than the ground state, due to the Boltzmann factor). Due to the increasing Stark broadening of these transitions from random electric fields as we approach the series limit, they merge and converge into a quasi-continuum well before the nominal series limit (at 3646 Å) is reached. As we approach the series limit this non-polarizing bound-bound opacity overtakes the polarizing Lyman scattering opacity, which leads to an effective Balmer jump in the degree of linear polarization before the series limit is reached.

Only a small fraction of the bound-bound absorption transitions represent scattering (spontaneous emission back to level $n = 2$ after the radiative absorption from level $n = 2$), but this opacity source can be neglected in comparison with the others, except near the center of each Balmer resonance.

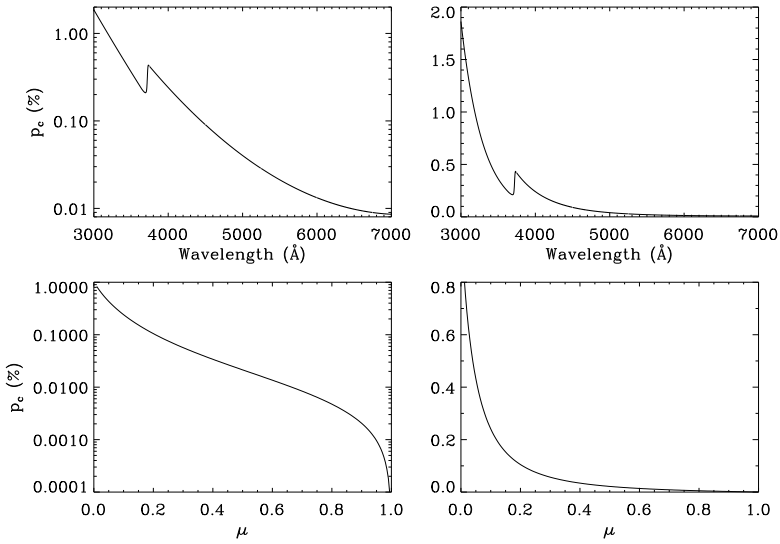


Fig. 3. Linear polarization of the Sun's continuous spectrum as a function of wavelength (top panels) and center-to-limb distance as parametrized by μ (bottom panels), from Stenflo (2005). The top panels are empirical determinations for $\mu = 0.1$, based on the Atlas of the Second Solar Spectrum (Gandorfer, 2000, 2002, 2005), while the bottom panels are for a wavelength of 4000 Å and based on a scaled version of the theoretical results of Fluri & Stenflo (1999).

Figure 3, from Stenflo (2005), illustrates how the polarization of the Sun's continuum varies with wavelength and μ (representing the center-to-limb distance). The left panels give the polarization in log scale, the right panels in linear scale. The wavelength dependence, given in the top panels for $\mu = 0.1$, shows a steep increase in the polarization as we go to shorter wavelengths. This increase has two sources: (a) The limb darkening, being the source of the anisotropy of the radiation field, increases strongly with decreasing wavelength. The polarization scales with the degree

of anisotropy. (b) The ratio between the polarizing Lyman scattering opacity and the non-polarizing H^- opacity increases with decreasing wavelength. The Balmer jump, which lies about 100 Å redwards of the nominal series limit due to the merging of the crowded bound-bound transitions, is produced when the non-polarizing Balmer absorption becomes larger than the Lyman scattering opacity (cf. Fig. 2).

The wavelength dependence in the two top panels is not based on any theory but is entirely empirical. It has been extracted from Gandorfer's Atlas of the Second Solar Spectrum (Gandorfer, 2000, 2002, 2005) through an elaborate reduction procedure (cf. Stenflo, 2005). The center-to-limb variation in the two bottom panels is however based on the radiative-transfer theory of Fluri & Stenflo (1999), scaled to make the $\mu = 0.1$ amplitude agree with the empirically determined values.

The center-to-limb variation, given in the bottom panels for a wavelength of 4000 Å, differs considerably from the corresponding Chandrasekhar solution in Fig. 1, in two main respects: (1) The center-to-limb variation of the continuum polarization is much steeper. (2) The polarization amplitudes are much smaller, by an order of magnitude near the limb, and much more at larger limb distances. There are several reasons for this difference: The limb darkening for a purely scattering atmosphere is very different from that of the Sun's atmosphere, which is strongly wavelength dependent. The ratio between the polarizing scattering opacity and the non-polarizing H^- opacity is of order 0.1 (although wavelength dependent) and varies with height. This ratio decreases with increasing depth in the atmosphere. Since we for larger μ look deeper into the atmosphere, below the main scattering layer, the polarization decreases faster with increasing μ than for the Chandrasekhar solution.

Not only the continuum but also the polarization inside resonance lines is found to exhibit a similar qualitative behavior, with a much steeper center-to-limb variation than the Chandrasekhar solution. The expected line polarization is determined largely by the height variation of the radiation-field anisotropy on the one hand (governed by radiative-transfer physics) and the intrinsic polarizability of the line transitions on the other hand (governed by the quantum structure of the scattering system).

3. Quantum physics and intrinsic polarization

The intrinsic polarizability of a scattering transition generally depends on the total angular momenta of the atomic levels that are involved in the scattering event. The polarizability can be characterized by the parameter W_2, which is the fraction of scattering events that behave like classical dipole scattering. The remaining fraction, $1 - W_2$, then behaves like isotropic, unpolarized scattering. Usually it is the J quantum numbers of the initial, intermediate, and final levels that are relevant, but in systems with non-zero nuclear spin I, this spin couples with J to form a new total angular momentum quantum number F, which then becomes the relevant one. However, these quantum numbers are only part of the story: the intrinsic polarizability can be drastically altered by quantum interferences between the atomic levels as well as by polarization transfers through optical pumping.

Figure 4 illustrates the effect of quantum interference between atomic states with different J quantum numbers. The top panel shows the intensity spectrum cover-

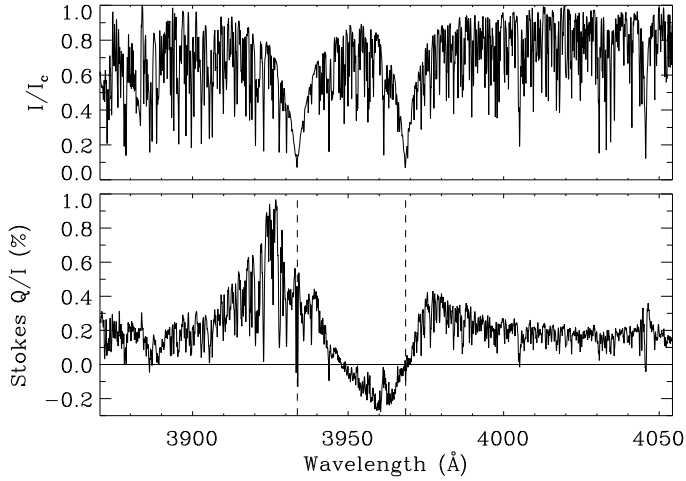


Fig. 4. Illustration of quantum interference in the Ca II K and H line transition system. The recording of the intensity (top panel) and degree of linear polarization (Stokes Q/I, bottom panel) over a range of 180 Å at disk position $\mu = 0.14$ was obtained in 1978 at the Kitt Peak Observatory (Stenflo, 1980). The dashed vertical lines in the bottom panel mark the wavelengths of the K and H line resonances.

ing a range of 180 Å around the K and H lines of singly ionized calcium (Ca II 3933 and 3968 Å, respectively), while the bottom panel gives the simultaneously recorded degree of linear polarization. The recording was made with the spectrograph slit placed 10 arcsec inside the solar limb (Stenflo, 1980). The myriad of spectral structures that we see are not due to noise but to the large number of blend lines in this spectral region. Here we focus our attention on the large-scale envelop variation, which is due to the Ca II K and H system. We notice that the polarization is highly asymmetric around the K line, becomes negative (meaning that the plane of polarization is perpendicular to the solar limb) for a long stretch between the two lines, and then has a zero-crossing at the center of the H line resonance. This strange behavior is governed by quantum interference.

The K and H line transitions have a common ground state, with $J = \frac{1}{2}$, while for the excited state $J = \frac{3}{2}$ for the K line and $\frac{1}{2}$ for the H line. With this quantum number combination the H line should be intrinsically unpolarizable ($W_2 = 0$), while the K line should have $W_2 = \frac{1}{2}$. The two transitions are however not independent of each other. When the calcium ion is radiatively excited, it does not choose between the $J = \frac{3}{2}$ and $\frac{1}{2}$ states, but the intermediate state is a mixed quantum state, a linear superposition of the two J states (Stenflo, 1980). The situation is fully analogous to the double-slit experiment, in which each photon has to pass through both slits at the same time, or to the Schrödinger cat, which is in a superposition of being both dead and alive. The level interference that results from this quantum superposition dramatically changes the intrinsic polarizabilities W_2 of the scattering system, in the same way as the interference pattern from a double slit is qualitatively different from an incoherent superposition of two single-slit interference patterns.

As the next example we show the polarization in the D_2 and D_1 lines of neutral

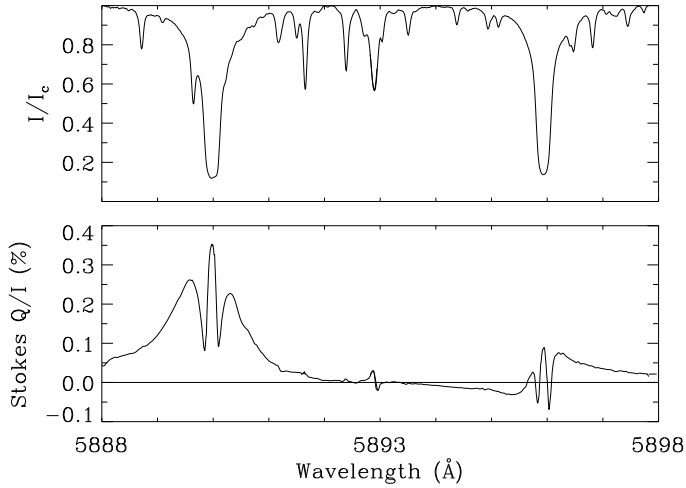


Fig. 5. Polarization structure from scattering at the sodium D_2 (5889.97 Å) and D_1 (5895.94 Å) system, as observed at 5 arcsec inside the limb of the quiet Sun (Stenflo *et al.*, 2000a). The polarization peak at the center of the D_1 line still remains an enigma.

sodium (at 5889.97 and 5895.94 Å), from (Stenflo *et al.*, 2000a). These line transitions have the same J quantum number combinations as the K and H transitions that we have just discussed, but with an important difference: In contrast to calcium, sodium has a nuclear spin, $I = \frac{3}{2}$, which gives rise to a set of new levels with different combinations of total angular momentum quantum numbers F. However, even with these new quantum numbers the D_1 line at 5895.94 Å should remain intrinsically unpolarizable ($W_2 = 0$ at the center of the line). In contrast, the observations reveal a small but significant polarization peak centered around the D_1 resonance.

Another mechanism that has considerable influence on the intrinsic line polarizability is polarization transfer by optical pumping, as suggested by Landi Degl'Innocenti (1998, 1999). The atomic polarization of the excited state that is induced by the anisotropic illumination is partially transfered and mapped into the otherwise unpolarized ground state through the spontaneous emission process. After many scattering processes a statistical equilibrium is reached, in which the ground state possesses atomic polarization. Scattering from a polarized ground state can produce greatly different scattering polarization as compared with scattering from an unpolarized state.

While such optical pumping has been found to play a decisive role in certain atomic multiplets (cf. Sect. 5 below), it is insufficient for explaining the observed D_2 – D_1 polarization, since the observed polarization amplitude is 1-2 orders of magnitude larger and the symmetry opposite to the theoretical prediction (Trujillo Bueno *et al.*, 2002; Kerkeni & Bommier, 2002; Casini *et al.*, 2002; Klement & Stenflo, 2003; Casini & Manso Sainz, 2005). The origin of this discrepancy has not yet been identified. To clarify the underlying physics we have therefore set up a laboratory experiment at ETH Zurich to study the polarizing properties of scattering at potassium vapor (Thalmann *et al.*, 2006). Potassium has the same quantum structure

as sodium and has been chosen for the experiment, since solid-state tunable lasers are available for the potassium wavelengths, in contrast to the wavelengths of the sodium lines. The experiment is still in progress.

4. Unique role of eclipse observations

We have seen in the previous sections that the observed scattering polarization on the Sun is much smaller than the values of the Chandrasekhar solution, but that the observed center-to-limb variation is much steeper. In the Chandrasekhar solution the opacity is determined exclusively by polarizing dipole-type scattering at all heights in the atmosphere, while on the real Sun scattering becomes insignificant deep in the atmosphere, where the collisional rates dominate over the radiative rates. As we go up in the atmosphere the density and therefore also the collision rates decrease until the radiative rates take over. Therefore the polarizing scattering layer lies systematically above the non-polarizing opacity layers. As we approach the limb we sample increasingly higher layers of the atmosphere, where polarizing scattering plays an increasingly significant role. At the top of the atmosphere, corresponding to the extreme limb where $\mu = 0$, it should be possible to disregard collisions entirely, which implies that we approach conditions that are much more similar to the pure scattering case considered by Chandrasekhar.

It should be noted, however, that this does not imply that the Chandrasekhar solution should be asymptotically reached in quantitative detail. Even if the top layer were governed exclusively by scattering processes, and even if the scattering particles would have intrinsic polarizability $W_2 = 1$ (like classical scattering), the polarization at the extreme limb would still differ from the Chandrasekhar case. The main reason is that the solar limb darkening, which determines the anisotropy of the radiation field that illuminates the scattering layer, is governed by non-local processes in deeper layers that are not dominated by scattering, in contrast to the idealized atmosphere governed by scattering at all heights. Still there is reason to expect that there should be a dramatic enhancement of the polarization amplitude within the last couple of arcsec of the extreme limb, such that polarization amplitudes in some spectral lines would asymptotically reach magnitudes comparable to the Chandrasekhar limit of 11.7 %. To deepen our general understanding of the formation of stellar spectra it would be of great interest to explore how the scattering polarization varies at the extreme limb in different parts of the spectrum, both through theoretical modelling and through observations. Neither has been done so far.

The reason why the polarization at the extreme limb has not been explored before is that one really needs a total solar eclipse to do it well. The window of opportunity occurs only during the flash phase, during the brief but dramatic transition between the partial and total phase at second and third contact of the eclipse sequence. The limb of the Moon serves as a knife edge that successively covers the different height layers of the Sun's atmosphere. Since the spectral intensity drops steeply with height, the contribution to the flash spectrum is dominated by the atmospheric layer that lies immediately adjacent to the Moon's limb. As the Moon moves about $\frac{1}{2}$ arcsec per second, we would obtain the phenomenal height resolution of order 10 km if we

use a frame rate of 50 frames/s, which is the nominal choice for our 29 March 2006 eclipse experiment (Feller *et al.*, 2006).

The two main obstacles preventing good recordings of the extreme limb polarization outside an eclipse are (1) spatial resolution, and (2) stray light. The best spatial resolution that can currently be achieved corresponds to about 100 km on the Sun, or an order of magnitude worse in height resolution than the eclipse would offer. The difference is actually much larger than this, since 100 km resolution has never been reached in combination with sensitive polarimetry, and much less in combination with large spectral coverage (our eclipse experiment has a simultaneous spectral coverage of about 5000 Å, cf. Feller *et al.*, 2006). These limitations are equally valid for both ground-based and space-based instruments. While the behavior of the spectral distribution of the scattering polarization (the "Second Solar Spectrum") has been extensively explored during the past decade for disk positions up to 5 arcsec inside the Sun's limb, it is the dramatic polarization increase within the last couple of arcsec that has not yet been captured, and for this we need a solar eclipse.

Due to the steep intensity drop at the extreme limb, small amounts of stray light in the instrument or in the Earth's atmosphere may seriously contaminate the observations. Instrumental stray light may be suppressed by a coronagraphic design, using an occulting disk to produce an artificial eclipse. However, scattering from the illuminated bright edge of the occulting disk will be a serious source of stray light (in addition to scattering at microscopic inhomogeneities in the optical components). Using optical tricks like the Lyot design with a field lense, the scattered light from the edge of the occulting disk can be significantly suppressed but not eliminated. The remaining stray light would still cause serious contamination. Note that in ordinary coronagraphs the occulting disk is significantly larger than the solar disk (in space coronagraphs like LASCO it is much larger), so that the edge of the occulting disk is not directly exposed to the solar disk. To isolate the different height layers at the extreme limb, however, the edge of the occulting disk has to be a fraction of an arcsec *inside* the limb and be shifted progressively outwards. The edge is therefore directly exposed to the bright disk, which makes the stray light problem much more serious than in standard coronagraphs. Note also that the telescopes that currently reach the best spatial resolution, approaching 100 km, do not have a coronagraphic design.

A method to remove most of the effects of image motions due to atmospheric seeing and thereby achieve a higher limb and μ resolution has been developed and implemented by Sheeley & Keller (2003) by recording a large number of frames with high time resolution and then using the intensity level of each frame to assign a μ value and organize the frames in μ bins. While this rather laborious method removes much of the seeing smearing due to image motions (without the use of adaptive optics), the limitations of stray light and telescope resolution are not overcome.

At an eclipse the Moon is our occulting disk. In contrast to artificial occulting disks both the spatial resolution and stray light limitations are entirely eliminated when the Moon is used. The reason why the spatial resolution problem is eliminated is that the exact position of the Moon's edge is fixed with respect to the Sun (for each given moment of time), entirely independent of any seeing, image motion, telescope pointing or resolution. The reason why stray light from the lunar occulting disk

is completely irrelevant is because the Moon is so far away, so that the solid angle occupied by the telescope aperture as seen from the Moon is vanishingly small. An artificial occulting disk in or near the telescope on the other hand will scatter into the instrument over a large solid angle.

The main imperfection of the Moon's edge as our knife edge is that it is not straight but somewhat jagged due to the lunar craters and valleys. This causes the Moon's limb to wiggle up and down by an amount that on the Sun may correspond to a few tens of km, depending on which part of the Moon's limb is used. With modest spatial resolution in the direction along the Moon's limb one may easily resolve these wiggles and account for them in the analysis, thereby avoiding significant reduction of the height resolution. Still the jaggedness of the Moon's limb will cause some moderate degradation, since local stray light from the brighter valleys may contaminate the fainter hills along the Moon's limb.

5. Differential effects in the polarized flash spectrum

We have seen in Sects. 2 and 3 that the polarized spectrum is highly structured and the expected polarization effects are very different in different spectral lines. The various line transitions have individual intrinsic physical properties and respond very differently to the atmospheric structure. These differential effects contain potentially rich diagnostic information about the solar atmosphere and the physics of line formation, but they also provide crucial help in the removal of instrumental effects in the data reduction process. In contrast the idealized Chandrasekhar solution has a universal center-to-limb variation without any spectral structures.

One of the notoriously most difficult problems in high-precision polarimetry is the determination of the true zero point of the polarization scale. The *absolute* polarization amplitudes generally cannot be determined with accuracies comparable to those of the *relative* amplitudes. When comparing two spectra in orthogonal polarization states, placed side by side (via a polarizing beam splitter), or recorded one after the other (via polarization modulation), their intensities will differ if polarization is present, but they may also differ for other reasons (imbalance of the beam splitter and of the detector gain table, or asymmetries in the modulation). These other effects will generate a fictitious polarization signal, usually in the form of an offset of the zero point of the polarization scale. To fix the zero point we need to lock it to some spectral feature for which we can assume that its intrinsic polarization must be zero or at least small in comparison with other spectral structures.

One example of a spectral feature with intrinsically zero polarization is the Ca II H line at 3968 Å, as we saw in Fig. 4. By locking the zero point of the polarization scale to this line and differentially relate the apparent Ca II K polarization to that of the H line, we can obtain the true, absolute value of the K polarization from the relative scale.

This differential procedure works best when comparing the relative polarization amplitudes for lines that belong to the same atomic multiplet and have different intrinsic polarizabilities, like the K and H line multiplet, or the sodium D_2 and D_1 multiplet (both of which only contain these two lines). Another very useful case is that of the Ca II infrared triplet at 8498, 8542, and 8662 Å, which is shown in Fig. 6,

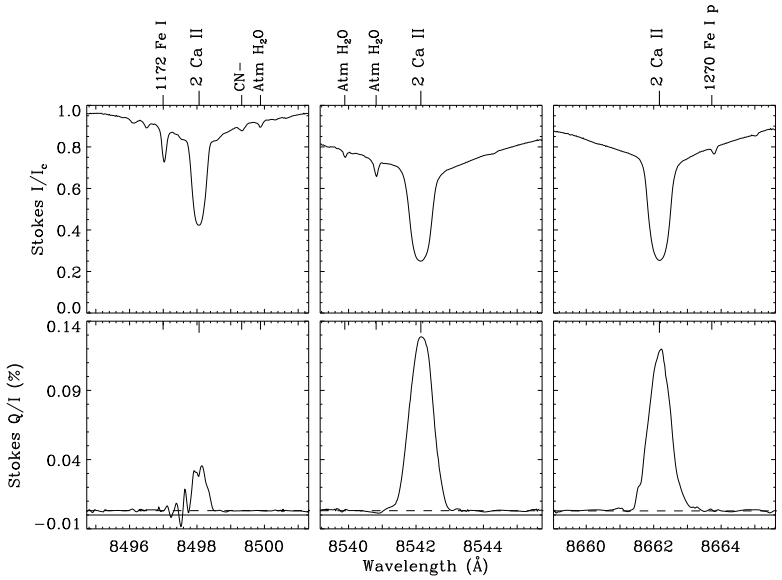


Fig. 6. Scattering polarization from the Ca II infrared triplet, as recorded at $\mu = 0.1$ (5 arcsec inside the solar limb) on the quiet Sun, from Stenflo *et al.* (2000b). The relative polarization amplitudes have been found to be governed by optical pumping (Manso Sainz & Trujillo Bueno, 2003).

from Stenflo *et al.* (2000b). This multiplet has the same upper levels as the K and H line multiplet, but has different lower states. It can therefore be fluorescently coupled to the K and H lines (radiative excitation in the ultraviolet, emission in the infrared).

According to the J quantum numbers of the transitions, the intrinsic polarizabilities W_2 would be expected to be zero for the 8662 Å line of the infrared triplet, while the 8498 Å line should be the most polarizable. The observed amplitude relations in Fig. 6 are however qualitatively entirely different, with the 8498 Å line exhibiting much less polarization than the 8662 Å line. An explanation for this behavior has been found in terms of optical pumping (Manso Sainz & Trujillo Bueno, 2003). The radiatively induced atomic polarization in the excited $J = \frac{3}{2}$ state is partially transferred to the lower states by the spontaneous emission process, until, after many such scattering processes, a statistical equilibrium is reached. Scattering from initially polarized lower states produces scattering polarization that is very different from the values expected from W_2 alone.

Since the optical pumping process is not a function of center-to-limb distance, one may use the observed ratios between the polarization amplitudes in the three lines as representative for the intrinsic ratios in the flash spectrum, and adopt this as a constraint to fix the location of the true zero point of the polarization scale in this infrared portion of the spectrum.

Another independent constraint on the true zero point is the qualitatively known approximate shape of the center-to-limb variation of the polarization: We know from

previous, non-eclipse observations that the polarization drops to values usually well below 1 % already a few arcsec inside the solar limb, and then continues to drop in a rather well defined way (cf. Fig. 3). The level of the zero point of the polarization scale has to be consistent with this observed fact.

6. Concluding remarks

With observations from space of the Sun's corona in EUV and X-rays, or with coronagraphs in the visible, the special circumstances offered by a total solar eclipse may appear to be less unique and scientifically attractive than they used to be in the past. There is however one scientific measurement that cannot be accomplished well without a total solar eclipse, namely the determination of the polarization of the flash spectrum. Such a measurement would give us the scattering polarization near the extreme limb of the Sun, as it is formed in the different height layers of the solar atmosphere. The Moon's limb serves as a knife edge that moves across the different height layers, covering them up or uncovering them in a well-defined way that is free from effects of seeing, telescope resolution, and stray light. The polarization of the flash spectrum is expected to have a structural richness that contains novel diagnostic information on the conditions in the solar atmosphere and on the physics of spectral line formation. Still the polarization of the flash spectrum has never been recorded with any success in the past.

The Chandrasekhar solution for an idealized, purely scattering atmosphere with a plane-parallel stratification exhibits very high polarization values, reaching the asymptotic value of 11.7 % at the extreme limb. Observations of the scattering polarization 5 arcsec or more inside the limb show much smaller values, generally much below 1 %, and the polarized spectrum is found to be extremely structured (and is therefore called the "Second Solar Spectrum"), in contrast to the Chandrasekhar solution. The main reason for this difference is that collisional effects and non-scattering opacities dominate in much of the solar photosphere. As we move up in the atmosphere, however, the collision rates decrease to become smaller than the radiative rates due to the drop in density with height. Scattering then plays an increasing role relative to true absorption. This implies that we approach conditions that are more similar to the idealized conditions that are assumed in the Chandrasekhar solution as we move up into the chromosphere. The steep increase in polarization that we already observe at a limb distance of 5 arcsec may therefore dramatically increase within the last arcseconds from the limb to possibly become comparable to the Chandrasekhar values. Whether this is really the case or not has never been explored before, but it needs a total solar eclipse to be empirically addressed.

During the flash phase, which has a duration of order 20 s, the Sun's spectrum changes dramatically from an absorption-line spectrum that originates in the photosphere to an emission-line spectrum that comes from the chromosphere. The precise location of the Sun's limb depends on wavelength and type of spectrum. The emission lines naturally originate above the photospheric limb. Therefore the translation of position into center-to-limb distance or μ becomes ambiguous and loses its meaning at the extreme limb or above.

While the assumption of a plane-parallel stratification has often worked well for modelling of the photosphere, it loses its validity when the line-of-sight has almost grazing incidence near the extreme limb, and due to the Sun's spherical geometry the horizontal optical thickness becomes finite instead of the semi-infinite plane-parallel stratification. The transition from the photosphere to the chromosphere is not only a transition from an absorption-line to an emission-line spectrum, but also a transition to a force-free, magnetically controlled plasma that is highly filamentary and dynamic, with large local fluctuations of the thermodynamic parameters. The resulting local fluctuations in the radiation-field anisotropy will affect the local values of the scattering polarization. Randomly oriented small-scall magnetic fields may cause depolarization via the Hanle effect. Realistic theoretical modelling of these effects is a difficult undertaking that still has to be done, but it needs observational guidance.

Ideally the observations during the flash phase should combine high polarimetric sensitivity with broad spectral coverage and high temporal, spectral, and spatial resolution. In the past nobody has ever recorded the flash spectrum successfully with high polarimetric sensitivity even without any of the other resolutions or spectral coverage. In our experiment for the 29 March 2006 eclipse at Waw an Namos we aim to record the flash spectrum with reasonably high polarimetric accuracy (promille range), large spectral coverage and high temporal resolution, but with modest spectral and spatial resolutions. It is a good beginning, but we have a long way to go. Improved generations of the experiment at future total solar eclipses will be needed to asymptotically approach the ideal of a perfectly resolved polarized flash spectrum.

References

Casini, R., Manso Sainz, R.: 2005, *ApJ*, **624**, 1025

Casini, R., Landi Degl'Innocenti, E., Landolfi, M., Trujillo Bueno, J.: 2002, *ApJ*, **573**, 864

Chandrasekhar, S.: 1950, *Radiative Transfer*, Clarendon Press, Oxford

Feller, A., Stenflo, J.O., Gisler, D., Ramelli, R.: 2006, these proceedings

Fluri, D.M., Stenflo, J.O.: 1999, *A&A*, **341**, 902

Gandorfer, A.: 2000, *The Second Solar Spectrum, Vol. I: 4625 Å to 6995 Å*, ISBN no. 3 7281 2764 7 (Zurich: VdF)

Gandorfer, A.: 2002, *The Second Solar Spectrum, Vol. II: 3910 Å to 4630 Å*, ISBN no. 3 7281 2855 4 (Zurich: VdF)

Gandorfer, A.: 2005, *The Second Solar Spectrum, Vol. III: 3160 Å to 3915 Å*, ISBN no. 3 7281 3018 4 (Zurich: VdF)

Kerkeni, B., Bommier, V.: 2002, *A&A*, **394**, 707

Klement, J., Stenflo, J.O.: 2003, in J. Trujillo Bueno and J. Sánchez Almeida (Eds.), *Solar Polarization*, Proc. 3rd SPW, *ASP Conf. Ser.*, **307**, 278

Landi Degl'Innocenti, E.: 1998, *Nature*, **392**, 256

Landi Degl'Innocenti, E.: 1999, in K.N. Nagendra and J.O. Stenflo (Eds.), *Solar Polarization*, Proc. 2nd SPW, *Astrophys. Space Sci. Library*, **243**, 61 (Dordrecht: Kluwer)

Manso Sainz, R., Trujillo Bueno, J.: 2003, *Phys. Rev. Let.* **91**, 111102

Sheeley, N.R., Jr., Keller, C.U.: 2003, *ApJ*, **594**, 1085

Stenflo, J.O.: 1980, *A&A*, **84**, 68

Stenflo, J.O.: 2004, *Rev. Mod. Astron.*, **17**, 269

Stenflo, J.O.: 2005, *A&A*, **429**, 713

Stenflo, J.O, Keller, C.U.: 1996, *Nature*, **382**, 588

Stenflo, J.O, Keller, C.U.: 1997, *A&A*, **321**, 927

Stenflo, J.O., Gandorfer, A., Keller, C.U.: 2000a, *A&A*, **355**, 781
Stenflo, J.O., Keller, C.U., Gandorfer, A.: 2000b, *A&A*, **355**, 789
Thalmann, C., Stenflo, J.O., Feller, A., Cacciani, A.: 2006, in R. Casini and B.W. Lites (Eds.), *Solar Polarization*, Proc. 4th SPW, *ASP Conf. Ser.*, in press
Trujillo Bueno, J., Casini, R., Landolfi, M., Landi Degl'Innocenti, E.: 2002, *ApJ*, **566**, L53

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

ECLIPSE INSTRUMENT TO RECORD THE POLARIZATION OF THE FLASH SPECTRUM

ALEX FELLER, JAN O. STENFLO * and DANIEL GISLER
Institute of Astronomy, ETH Zurich, 8092 Zurich, Switzerland

and

RENZO RAMELLI
Istituto Ricerche Solari Locarno, Via Patocchi, CH-6605, Locarno, Switzerland

Abstract. We have designed and built an instrument to capture the polarization of the flash spectrum during a total solar eclipse. The whole visible spectral range, from UV to the near IR, is covered with a spectral resolution of 3000. The sampling period is about 20 ms, which corresponds to a height resolution of some 10 km in the solar atmosphere. The instrument is composed of an 8-inch Dall-Kirkham type telescope and a slitless spectropolarimeter. The F/11.5 telescope beam projects a solar image in its focal plane, where the entrance aperture of the spectropolarimeter cuts a 2.5 arcmin slice of the solar limb. The resulting beam is split into two orthogonally polarized components by means of a Savart plate. Before entering the spectrograph the beam is reduced to F/2.8 to have sufficient illumination on the detector. The measurements are only feasible during an eclipse, since seeing and stray light prohibit a clean isolation of the chromospheric light with a coronagraph. The flash spectrum has been successfully recorded during the eclipse of March 29, 2006.

Key words: polarization, solar eclipse, flash spectrum

1. Introduction

The chromospheric emission spectrum, called flash spectrum, has already been recorded successfully during multiple eclipse expeditions (Cillié and Menzel, 1935; Mitchell, 1947; Dunn *et al.*, 1968; Shen *et al.*, 1981). The mentioned authors have compiled catalogues itemizing the calibrated absolute fluxes of more than a thousand emission lines in the range 300 to 910 nm. The spectra of Dunn *et al.* (1968) cover 4500 km on the Sun with a height resolution of 100 km in the direction of lunar movement. Although these data lack polarization information, they have been very helpful for the design of our instrument, allowing us to narrow down the required sensitivity range.

In terms of polarimetry we know of two past eclipse observations. The measurements in the Ca II H and K lines of Hanson *et al.* (1976) allowed a 5% upper limit to be set on the K line degree of polarization. Moreover, from his visual tracking of the intensity variations produced by a rotating polaroid screen, Stokley (1948) estimated that the overall degree of polarization of the flash spectrum is probably not above 10%. This result is however questionable because of the rapid exponential drop of the line intensities during the flash phase.

* also affiliated with the Faculty of Mathematics & Science, University of Zurich

The flash spectrum is extremely difficult to observe outside an eclipse. The steep intensity decrease at the extreme limb of the Sun makes a coronagraph very vulnerable to stray light caused by the earth's atmosphere and by the optics of the instrument itself. The main challenge here is the occulting disc. It must not exceed the solar disc by more than a fraction of an arcsecond, to properly isolate the thin chromospheric layer from the photosphere while still leaving it uncovered. Another serious drawback of out-of-eclipse observations is the spatial resolution of current solar telescopes, which is limited to about 100 km. In contrast, as the eclipsing Moon is moving at a relative speed of about 350 km/s, a height resolution of order 10 km on the Sun can already be achieved with a moderate frame rate of 50 frames/s.

When using the Moon as occulting disc one has to take into account that the lunar limb is somewhat serrated. For the 29 March 2006 eclipse, the lunar limb profile at the positions of second and third contact (Espenak and Anderson, 2004) shows typical height variations corresponding to about 700 km on the Sun. On the other hand, the lunar mountains and valleys can be easily resolved with a modest spatial resolution of about $5''$ in the direction parallel to the Moon's limb, which allows for a later correction of the chromospheric height scale across the field of view.

The scientific rationale of our observations is described in detail by Stenflo (2006). A theoretical reference has been developed by Chandrasekhar (1950) in terms of an idealized model for a plane-parallel purely scattering atmosphere. The conditions at the extreme limb of the Sun are approaching this Chandrasekhar limit, but the real polarization can nevertheless be significantly influenced by different physical processes: the relative importance of different opacities, collisions, deviation from plane-parallel stratification (spherical geometry at the limb, small scale inhomogeneities), atomic physics (quantum interference, optical pumping), radiative-transfer physics and magnetic fields (Hanle effect). These processes affect the individual spectral lines in different ways. By recording a large part of the spectrum, we may be able to untangle and quantify them with the help of differential diagnostics.

2. Instrument design

TABLE I
Main instrument characteristics.

spectral range	340 - 870 nm
spectral sampling	0.21 nm/pixel
time resolution	13 - 40 ms
spatial sampling	$5''$/pixel
field of view	$150''$

2.1. Photon budget

When designing a polarimeter one has to make a trade-off between spatial, spectral, and time resolution and polarimetric sensitivity (Table I). The chromospheric sickle is not resolved in practice in the direction of lunar movement, because of the steep

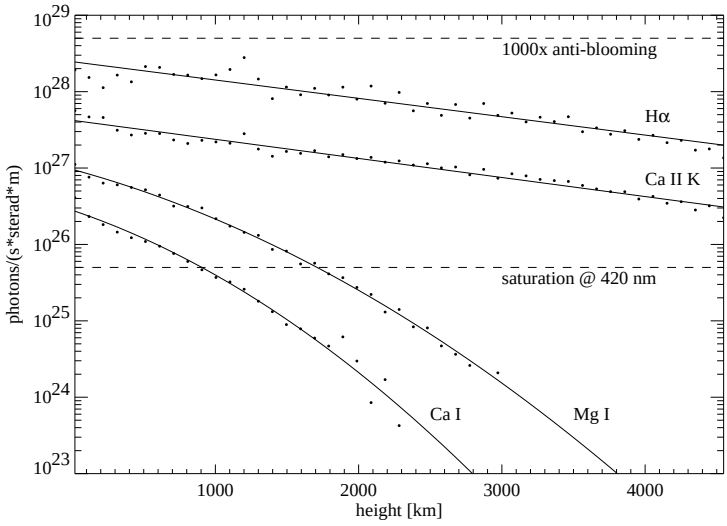


Fig. 1. Flux variation with height for some selected emission lines of Table II. Data are from Dunn *et al.* (1968). For our purpose, the logarithm of the flux can be sufficiently well fitted with a linear function for the strong lines and a second order polynomial for the weak lines. The estimated saturation level around Ca I and the corresponding anti-blooming limit are indicated by the dashed lines.

TABLE II
Lines of interest sorted by flux

Line	Wavelength nm	flux at h=0† photons · s^{-1} · $sterrad^{-1}$ · m^{-1}
Hα	656.3	$2.0 \cdot 10^{28}$
Ca II	854.2	$9.4 \cdot 10^{27}$
Ca II	866.2	$8.7 \cdot 10^{27}$
Ca II K	393.3	$5.5 \cdot 10^{27}$
Hβ	486.1	$5.4 \cdot 10^{27}$
Ca II H	396.8	$4.8 \cdot 10^{27}$
Ca II	849.8	$4.3 \cdot 10^{27}$
He D_3	587.6	$3.5 \cdot 10^{27}$
Mg I	518.4	$1.3 \cdot 10^{27}$
Na I D_2	589.0	$8.9 \cdot 10^{26}$
Na I D_1	589.6	$8.9 \cdot 10^{26}$
Mg I	516.7	$7.8 \cdot 10^{26}$
Mg I	517.2	$7.8 \cdot 10^{26}$
Ca I	422.7	$4.3 \cdot 10^{26}$

† from Dunn *et al.* (1968)

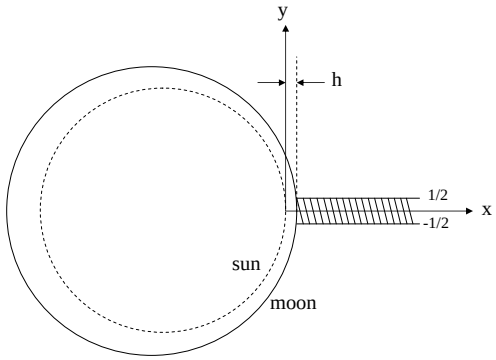


Fig. 2. Definition of the integrated flux from a chromospheric emission line.

intensity drop with height (see Fig. 1). The spectral line profile is not resolved either. In this context it is convenient to define, for a given line i, an integrated flux F_i given by

$$F_i(h) = \int_h^\infty \int_{-1/2}^{1/2} \int_{\lambda_i - \Delta\lambda_i/2}^{\lambda_i + \Delta\lambda_i/2} f(\lambda)\, d\lambda\, dy\, dx \tag{1}$$

Thus $F_i(h)$ is the total number of photons per second and sterad, emitted from a one unit thick semi-infinite slab of the chromosphere at height h (see Fig. 2) and within a spectral window $\Delta\lambda_i$ around the line center λ_i.

Given the flux F_i, we can then determine the total number of photoelectrons generated in a camera pixel.

$$N_i(h) = F_i(h)\, \Delta\Omega\, \Delta y\ \Delta t\, \epsilon_{\rm opt}\, q_{\rm eff}(\lambda_i) \tag{2}$$

$\Delta\Omega$ is the solid angle covered by the instrument aperture, as seen from the Sun, Δy is the extent along the chromospheric sickle imaged by the pixel, and Δt is the exposure time. $\epsilon_{\rm opt}$ refers to the total transmission of the atmosphere and the optics and $q_{\rm eff}$ to the quantum efficiency of the camera. As the product of atmospheric extinction and quantum efficiency is changing by a factor of 5 across the spectral working range, its wavelength dependence must be taken into account. The optics are roughly estimated to contribute to the total efficiency with a factor of 0.1.

The huge dynamic range between Hα and the weak lines (see Fig. 1 and Table II) cannot be handled by the camera at once. At maximum frame rate the instrument has to be sensitive enough to just saturate in the Ca I line at the beginning of the flash phase. The stronger lines will then drop into the sensitivity range later on. Unfortunately the Hα flux indicated by Dunn (1968) for $h = 0$ is not consistent with the flux at the center of the Hα absorption line at the photospheric limb, which is an order of magnitude lower. This uncertainty is however not critical for the optical design as an instrument that is too fast can easily be dimmed. In addition the flux in the weak lines is decreasing steeply enough to stop saturating the camera after the first few seconds into the flash phase. On the other hand it is important to

TABLE III
Camera characteristics

CCD size	4096 pixels × 96 pixels
pixel size	13 μm × 13 μm
dark current at 30 °C	$1.23 \cdot 10^3$ e^-/s
readout noise	70 e^-
sampling	50 e^-/count, 10 bit
readout time	6 ms/frame
frame rate†	1.53 - 75 frames/s
anti-blooming factor	1000

† limited by the mechanical shutter

have a CCD with an anti-blooming feature of at least 1000 × saturation, to avoid any crosstalk from the strong lines during the saturated phase. As we were able to obtain a very fast F/2.8 spectrograph, we can adequately expose the camera with a modest and easily transportable 8-inch telescope (see Fig. 3).

The noise in the measured degree of polarization Q/I is, in case of a beam splitting polarimeter,

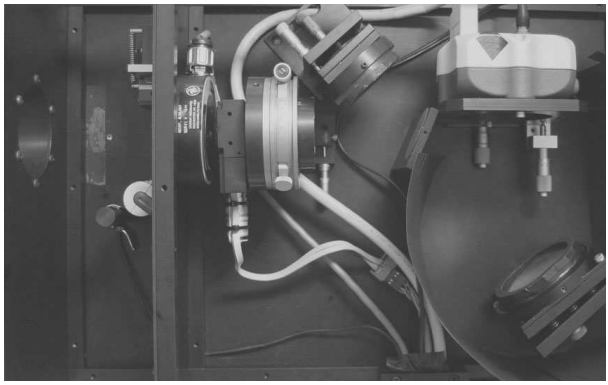
$$\sigma \approx \sqrt{\frac{1}{N} + \frac{N_{\rm dc}}{N^2} + \frac{\sigma^2_{\rm readout}}{N^2}} \quad , \tag{3}$$

where N is the number of photoelectrons, $N_{\rm dc}$ the number of electrons generated by the dark current and $\sigma_{\rm readout}$ the readout noise. All noise sources are assumed to be Poisson distributed. Close to saturation, σ is dominated by the shot noise $N^{-1/2} \approx 0.004$. This meets the requirements on polarimetric precision and is adequately sampled with a 10 bit A/D conversion (*cf.* Table III). At less than 0.1 × saturation the readout noise begins to dominate. $N_{\rm dc}$ stays below 0.001 × saturation and it's contribution to the polarimetric noise is negligible.

2.2. Telescope and spectropolarimeter unit

The telescope is an 8-inch Dall-Kirkham Cassegrain reflector. The imaging is nearly diffraction limited within our small field of view and the Cassegrain design is virtually free of instrumental polarization.

A scaled drawing and a close-up photograph of the spectropolarimeter unit are shown in Fig. 3. To keep the optical efficiency high and chromatic aberrations at a minimum, the use of glass in the beam path is avoided, except for the half wave plate and the beam splitter. At the telescope focus one can choose between a 50 μm slit of length 1.6 mm and an aperture of 1.6 mm × 1.6 mm. With the slit we record calibration spectra of the solar disk or a spectral lamp. The aperture is used for the actual observation of the flash spectrum. This slitless mode has some important advantages: it allows the imaging of extended solar structures and relaxes the tolerances on guiding and pointing of the telescope. Then again we observed disadvantages like aberrations caused by an off-axis position of the sickle or the smearing of nearby spectral lines.



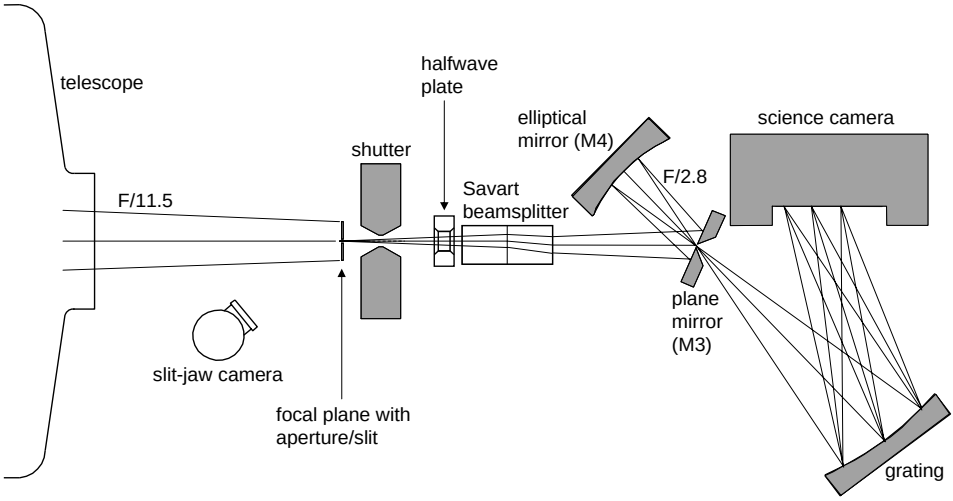


Fig. 3. **Bottom:** optical setup. **Top left:** top view of the spectropolarimeter unit. **Top right:** overview of the setup at the eclipse site near Waw an Namos, Libya. The spectropolarimeter unit is mounted behind the telescope. In the background one can see the tent which is used as work place and protection for the electronic equipment. The electrical power of the whole setup (150-200 W) is provided by fuel generators and secured with a UPS.

In terms of polarimetry we opt for a beam splitter, as opposed to the more accurate modulation/demodulation principle normally used in our Zurich Imaging Polarimeter (Povel, 1995; Gandorfer *et al.*, 2004). The arguments in favor of a beam splitter are the intricacies involved in developing an adequate achromatic modulator, and the low frame rate of our ZIMPOL cameras ($\sim$ 1 frame/s). Therefore a Savart beam splitter is used. It is installed in the slower F/11.5 beam to keep its aberrations below the size of a pixel. The disadvantage of this position is the need of rather thick calcite elements. We have 2×22.8 mm, giving us a beam separation between 3.40 and 3.76 mm, depending on wavelength. To eliminate the effect of the polarization dependent grating efficiency, the beam splitter has to be turned so that the polarization direction of both beams is oriented $\pm 45°$ to the grooves of the grating. A half wave plate is needed in this context to realign the Moon's limb with the

direction of Stokes $+Q$. We renounce the use of true beam exchange again because of the intricacies of getting an adequate fast-switchable and achromatic half wave plate. Instead the retarder is turned mechanically by 45° one single time between the two flash phases of second and third contact.

After the Savart plate the two beams are deflected by a plane mirror (M3) with a circular hole of 2.5 mm diameter, reduced by an elliptical mirror (M4) to F/2.8 and then focused through the hole in M3 into the spectrograph. This configuration allows to use M4 on-axis, and the hole only causes a small loss of about 5% of the flux. The focal points of M4 slightly deviate from their paraxial positions (telescope focus and intermediary focus at M3) to minimize the differential aberrations between cospatial image points in the two beams.

The spectrograph consists of an aberration corrected holographic concave grating with 405 grooves/mm, serving as both the dispersive and focussing element. Operated in first order it provides a practically linear dispersion of 0.21 nm/pixel over the whole wavelength range. The efficiency varies between about 45% (393 nm) and 23% (866 nm). We find however a significant residual astigmatism, clearly dominating the total aberrations of the instrument. As a trade-off we adjust the grating and camera position to align the detector plane as good as possible with the spectral focal plane, but accepting therefore a strong wavelength dependent spatial defocussing.

Both the telescope and the spectropolarimeter unit are assembled on a parallactic mounting with computer-controlled stepping motors on both axes (*cf.* Sect. 2.3). The guiding and pointing accuracies in the hour axis are limited by the worm gear which has a period of about 12 minutes at guiding speed and a mean amplitude of some $40''$. Due to their periodicity, the hour axis errors can at least be partly compensated with the guiding software. The guiding and pointing accuracies in the declination axis are of order $1''$/minute and $10''$ respectively and can be neglected for our purpose. The knowledge of the pointing accuracies is crucial for the eclipse observations as the telescope has to be moved blindly to the opposite limb during totality (*cf.* Sect. 2.3). It defines the minimum size of the aperture in the spectropolarimeter unit.

2.3. Instrument control and observing procedure

Figure 4 shows a block diagram of the different control units and programs. The control tasks are allocated to three different computers: Computer I is responsible for the guiding and acts as time server, computer II is used for monitoring the solar disc, and computer III is handling the spectropolarimeter unit.

The telescope mounting is aligned by star tracking according to the Bigourdan procedure. For accurate pointing on the Sun a special guiding software has been developed in-house which uses the solar limb as reference frame. The pointing position is determined by scanning the solar disc in the two axis directions with the spectropolarimeter in slit-mode, and determining the limb position by the corresponding inflection point in the measured intensity.

The time server is providing UTC with an accuracy of about 10 ms, which corresponds to the time resolution of the instrument.

The science camera and shutter as well as the rotation of the half wave plate

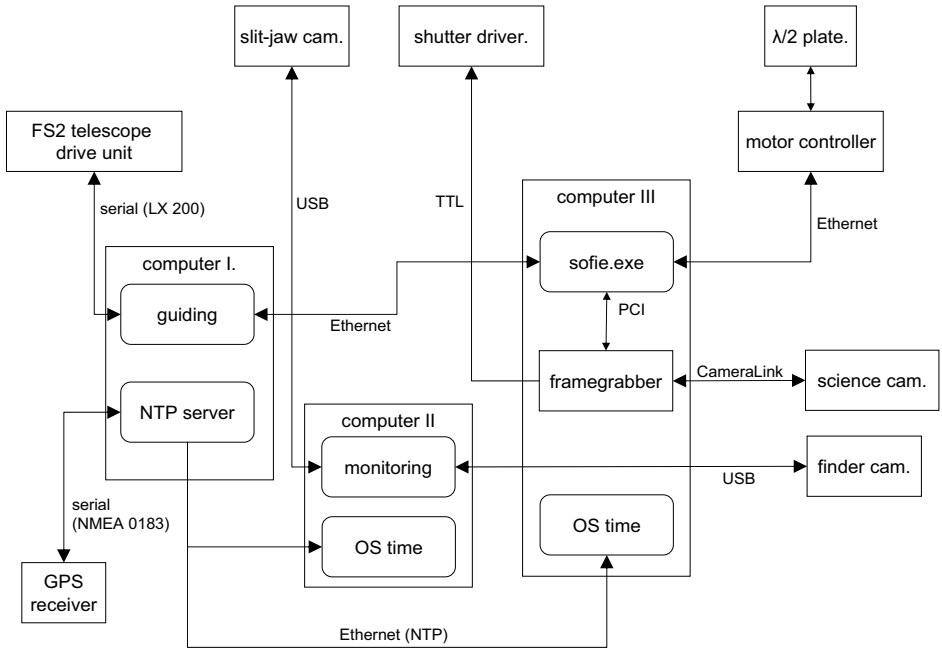


Fig. 4. Control units and programs for the eclipse instrument. The boxes with the soft edges denote computer programs, the boxes with sharp edges hardware devices. The lines of communication are annotated with the connection type and/or the protocol.

are controlled with software developed in-house and optimized for the particular requirements of the eclipse observation. A critical item is the reliable handling of the data stream, which is reaching 56 MB/s at maximum frame rate. Other important features are the live display of the science camera images and the possibility to quickly adjust the frame rate during operation.

TABLE IV
Location

Latitude	24:28:03 N
Longitude	17:57:52 E
Altitude	450 m

Table V summarizes all the relevant observing and calibration procedures together with the respective timings. Our observing spot is located near Waw an Namos in the Sahara of southern Libya (*cf.* Table IV), on the center line of the eclipse path and close to the point of maximum duration of totality. The center-line position is necessary to avoid sickle movements during the flash phases and it has the advantage the sickles of second and third contact are parallel.

TABLE V
Observing procedure

UTC	Events
	Limb scan to set the guiding reference frame
	Position the slit at the predicted position of first contact (angle 135.2°)
08:55:54.67	First contact
	Verify the positioning accuracy and the orientation of the spectropolarimeter unit
	Move to the predicted position of second contact (angle -44.7°)
	Limb scan including the lunar limb to correct the guiding reference frame
	Change from slit to aperture
10:13:30	Open the telescope
	Start recording with a fixed rate of 25 frames/s
10:13:45.20	Second contact
10:14:34	Slow down to 1.5 frames/s for the coronal lines
	Move to predicted position of third contact (angle 136.9°)
	Rotate half wave plate by 45° for beam exchange
10:17:14	Start with variable frame rate program: the frame rates are incremented from 25 to 75 frames/s in steps of 25 frames/s, based upon a visual rating of the flash intensity
10:17:55.74	Third contact
10:18:16	Close telescope
	Stop recording
	Change back to slit
	Polarimetric calibration at disc center
	Dark current/bias and flat field measurements
	Spectral calibration with Rb I lamp

3. Conclusions and outlook

The instrument described here has been developed and built from scratch within one year for the 29 March 2006 eclipse. The observing location in the desert, hundreds of km away from any infrastructure, called for robust, compact, reliable and completely autonomous equipment as well as elaborate logistics.

Despite all these challenges the equipment was running faultlessly. The eclipse was passing under ideal weather conditions and we were able to measure both flash phases of second and third contacts, on opposite solar limb positions (*cf.* Table V).

Our requirements were however not entirely satisfied. The main drawback of the instrument is the residual aberrations caused by the grating, in particular the differential effects between the two beams (*cf.* Fig. 5). They have been discovered only during the tests of the final instrument, and we were not able to fully resolve them in due time. On the other hand the slitless mode has proven to be an advantage for isolated spectral lines like Hα, allowing to image extended structures like a prominence, but a disadvantage for adjacent lines like He D_3, Na D_1 and D_2.

The data reduction is still ongoing, and the results as well as the reduction

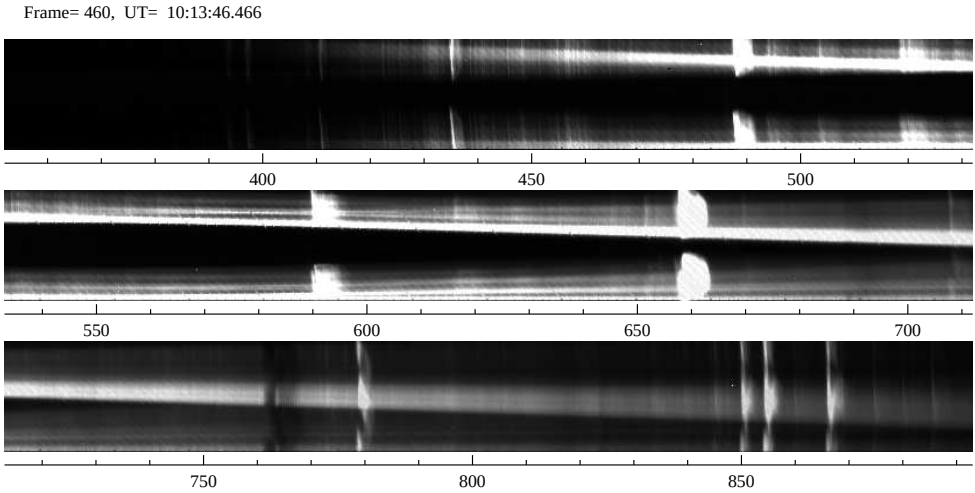


Fig. 5. Sample frame from the flash phase at second contact, representing the raw data (only corrected for bias and flat field). The scale gives the wavelength in nm. The two orthogonally polarized spectra are imaged above each other. Since the separation between them is strongly wavelength dependent, they are well separated in the UV (so well that vignetting is severe), while they significantly overlap in the infrared. While some strong lines, like Hβ, the Helium D3 lines, and Hα, are still saturated in this particular frame, the Ca H and K lines in the UV are already quite faint. The almost horizontal bright band represents the remaining part of the photospheric sickle. In this frame one can discern the main instrumental problems: vignetting, overlap, and aberrations, in particular the differential effects between the two spectra.

techniques will be discussed in detail in a later publication. The main challenge is to study and model the aberrations in order to determine the cospatial regions in both images and to improve the spatial resolution in some parts of the spectrum.

Finally we are already thinking of incorporating our experiences for a second generation eclipse experiment. The key enhancements we will be working on are the replacement of the slitless mode with an array of optical fibers, the use of an Echelle grating, and a true achromatic beam exchange.

Acknowledgements. Financial support for this work has been obtained through grants Nos. 200020-101603 and 200020-109159 from Swiss Nationalfonds. We would like to thank the technical staff at the Institute of Astronomy of ETH Zurich as well as Michele Bianda and Evio Tognini from the Istituto Ricerche Solari Locarno for helping with the development of the instrument and the control software and for handling the time pressure so calmly. We are also grateful to the Specola Solare Ticinese in Locarno, for having made available to us a very convenient testing environment, and to Osama Shalabiea and the Sebha University for their kind support in Libya.

References

Chandrasekhar, S.: 1950, *Radiative Transfer*, Clarendon Press, Oxford
Cillié, G.G. and Menzel, D.H.: 1935, *Harvard College Obs. Circ.*, **410**, 1
del Toro Iniesta, J.C. and Collados, M.: 2000, *Appl. Opt.* **39**, 1637
Dunn, R.B., Evans, J.W., Jefferies, J.T., Orrall, F.Q., White, O.R., and Zirker, J.B.: 1968, *ApJS*, **15**, 275
Espenak, F. and Anderson, J.: 2004, *NASA Technical Paper No.* **212762**

Gandorfer, A. M. et al.: 2004, *A&A*, **422**, 703
Hanson, T., Stenholm, B., Söderhjelm, S.: 1976, *Reports from the Observatory of Lund, No.* **8**
Mitchell, S.A.: 1947, *ApJ*, **105**, 1
Povel, H.P.: 1995, *Opt. Eng.*, **34**, 1870
Shen L., Li, Q., Yu, J., Sun, R.: 1981, *Chin. A&A*, **5**, 388
Stenflo, J.O.: 2006, these proceedings
Stokley, J.: 1948, *AJ*, **53**, 117

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

USING POLARIMETRIC IMAGING AND SPECTROSCOPY OF THE CORONA FROM 400 TO 1800 NM FOR EXPLORING THE NEAR SUN PLASMA

SHADIA RIFAI HABBAL, JEFF KUHN, DONALD MICKEY, SARAH JAEGGLI, HUW MORGAN and ILIA ROUSSEV
Institute for Astronomy, University of Hawaii, Honolulu, HI 96822, USA

JUDD JOHNSON
Electricon, Boulder, CO 80204, USA

MARTINA BELZ ARNDT
Department of Physics, Bridgewater State College, Bridgewater, MA 02325, USA

ADRIAN DAW
Department of Physics and Astronomy, Appalachian State University, Boone, NC 28608, USA

and

MUNIR H. NAYFEH
Department of Physics, University of Illinois at Urbana-Champaign, IL 61801, USA

Abstract. Total solar eclipses offer unique opportunities for exploring the solar corona, in particular for validating new concepts, and testing new instrumentation. The scientific goals of the experiments, planned for the total solar eclipse of 29 March 2006, at Waw AnNamous, Libya, by the eclipse group from the Institute for Astronomy (IfA) at the University of Hawaii, were as follows: (1) to search for the signature of silicon nanometer size dust grains in the corona, (2) to search for the signature of a neutral wind of interstellar origin in the corona, (3) to understand the nature of near-Sun dust from F-coronal visible and IR spectra, and (4) to establish the direction of the coronal magnetic field from its origin at the Sun as it expands into interplanetary space. A complement of imaging and spectroscopic polarization measurements, covering the wavelength range from 400 to 1800 nm, were designed and constructed to achieve these objectives. The observing conditions during the total solar eclipse deemed ideal, and the experiments were successfully executed to achieve these milestones.

Key words: solar corona, infrared emission, polarization, interplanetary dust, magnetic fields

1. Introduction

A number of novel concepts regarding the composition of the solar corona have recently emerged from coronal observations made in the near-infrared part of the solar spectrum during total solar eclipses.

In their eclipse observations of 21 June 2001, Habbal *et al.* (2003) found a tangentially polarized emission within the 0.5 nm bandpass of the Fe XIII 1074.7 nm filter. This errant polarization appeared in low-temperature and low-density regions of the corona, such as coronal holes, where this emission should have been absent. This was in contrast to the predominantly radial polarization direction observed in

the rest of the corona. This tangential polarization could not be accounted for by the theory of radiative excitation of coronal forbidden lines, such as the Fe XIII line (Charvin, 1965). In addition, the radial profile of the Fe XIII intensity as a function of radial distance, in the coronal regions with tangential polarization, dropped sharply close to the Sun, as expected from collisional excitation, but then flattened considerably when compared to the emission in the continuum (see Figure 1a).

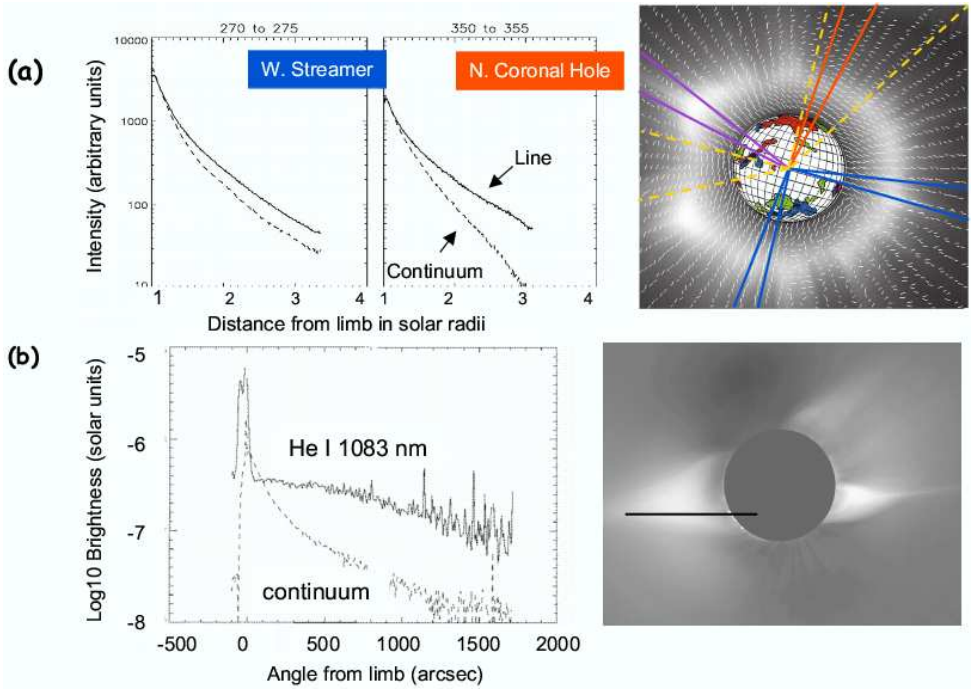


Fig. 1. Composite from the eclipse observations of (a) 2001 by Habbal et al. (2003) and (b) 1994 by Kuhn et al. (1996). (a) Radial profiles of the intensity of the Fe XIII 1074.7 nm line and neighboring continuum in arbitrary count units, versus radial distance, within 5 deg swaths given by the position angles above each panel. (The position angle is measured counter-clockwise from 0 degrees north). To the right is a map of the polarization angle for the 1074.7 nm Fe XIII line superimposed on the white-light image of the corona taken at the same time. The thick lines outline different regions in the corona where radial profiles of the intensity were plotted. (b) Radial profile of the He I 1083 nm line intensity compared to the continuum. To its right is the white light image of the corona with the spectrograph slit position and extent given by the straight line.

Taken together, the unexpected polarization direction and radial fall-off of line intensity were attributed to fluorescence from silicon nanometer dust grains in the inner corona. The assertion was based on observations in the interstellar medium (see, e.g. Witt *et al.*, 1998; Nayfeh *et al.*, 2005,) and on recent laboratory experiments by M. Nayfeh and his group (Nayfeh *et al.*, 2001; Belomoin *et al.*, 2002) who discovered silicon nanoparticle distributions in discrete and reproducible family of sizes, with diameters of 1, 1.67, 2.15, 2.9 and 3.7 nm. These particles become highly fluorescent in the visible, red, and infrared part of the spectrum under UV radiation. If present in the corona, the observed tangential polarization could be an indication

that this absorption/fluorescence process is a consequence of changes in the internal structure of the nanoparticles as a result of vibrational and rotational modes triggered by UV excitation. This conjecture is further supported by model calculations (Mann *et al.*, 2000) indicating that the radiation pressure force for silicates 5–10 nm in size is smaller than the gravitational force, thus enabling nanometer-size dust grains to get very close to the Sun.

During the 1994 eclipse, Kuhn *et al.* (1996) reported the discovery and confirmation of the existence of a number of spectral lines in the near infrared part of the spectrum. In particular, the S IX emission at 1252.5 nm was detected, and the Si X emission at 1430 nm was confirmed. In addition, an extended emission was observed in the He I 1083 nm line of neutral helium, with a redshift increasing with heliocentric distance to approximately 20 km/s. As shown in the reproduction of Figure 1b, the slit of the spectrograph in that experiment coincided with a high density streamer. Close to the Sun, one end of the slit overlapped a prominence, and the slit extended to about 3 R_s outwards. Curiously, the decrease in the He I line intensity with radial distance resembled the behavior of the Fe XIII line radial profile in a coronal hole in the 2001 eclipse measurements, characterized by a sharp drop off very close to the Sun, followed by a shallow profile. Hence, in both cases, the unexpected shape of the radial profiles occurred in coronal regions where emission from either Fe XIII (2001 data) or He I (1994 data) should have been absent.

Recently, the He I emission in 1994 data has been attributed to the focussing of the interstellar neutral He wind in the corona by solar gravity. This interpretation was supported by recent observations made on 14 February 2005 (Kuhn et al. 2006) with the coronagraph on Haleakala (SOLARC), and by ultraviolet coronagraph observations in the extended corona by the UVCS instrument on SoHO (e.g., Michels *et al.*, 2002).

Given the importance of the near-infrared part of the spectrum, it behooved us to focus on this wavelength range to follow up on these novel earlier findings, and to exploit the diagnostic power of polarimetric imaging and spectroscopy of the corona during the total solar eclipse of 29 March 2006 at Waw AnNamous, Libya. In what follows we briefly describe the suite of eclipse experiments (section 2), and present preliminary results (section 3). We conclude in section 4.

2. The IfA/UH eclipse experiments

A suite of three sets of instruments were designed and constructed to carry out polarimetric imaging and spectroscopy of the corona from 400 to 1800 nm.

2.1. The Near-Infared Imaging Spectropolarimeter

The Near-Infrared Imaging Spectropolarimeter (NIIS) was designed for imaging spectroscopy of the corona, with coarse spatial resolution but relatively high spectral resolution, by measuring intensities of several coronal emission lines simultaneously. The instrument comprises five major optical components: a simple telescope, a linear polarizer, an optical fiber bundle, a grating spectrograph and a camera.

The telescope optics consists simply of a spherical mirror of 7.5 cm diameter, tilted to form an image of the corona to the side of the telescope entrance aperture.

At its focus is a hexagonal array of 127 optical fibers, spaced uniformly to sample the corona out to 4 R_s, with a spatial resolution of about 7.5 arc minutes. The telescope is defocused so that each fiber samples a region about equal to the spacing between fibers.

Placed in front of the fiber array is a linear polarizer which may be rotated under computer control. The polarizer was rotated 45° between exposures, so the polarization of the coronal light can be obtained from a combination of four exposures.

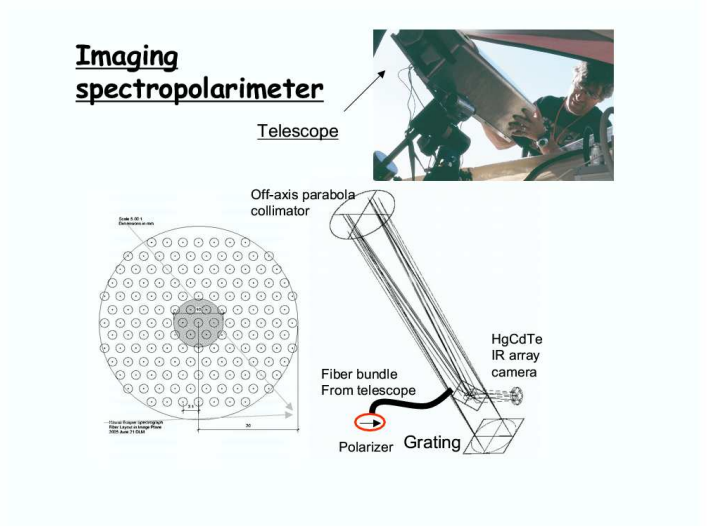


Fig. 2. The near-infrared imaging spectropolarimeter design, and a photo of the telescope part as mounted in the observing tent.

The optical fibers are rearranged at their output end into a linear array, with the fibers packed closely together. This linear array becomes the effective entrance slit of a Littrow-configuration reflecting grating spectrograph. The spectrograph produces a separate 2-pixel FWHM-wide spectrum from each fiber, with a resolving power of approximately 500. A 1024×1024 liquid nitrogen-cooled HgCdTe array camera is then used to record all 127 spectra simultaneously. The array was manufactured by Raytheon, Inc., and to our knowledge, is the first time these novel HgCdTe detectors have been used for solar astrophysics. A short-pass filter was used to limit thermal background at the detector and results in a useful wavelength range from 1 to 1.9 μm (See Figure 2).

2.2. The integrated visible wavelength spectrometer

A fiber optic spectrograph system was designed and built to measure the spatially integrated spectrum of a 2.5 R_s diameter annular region of the corona from 400 to 1100 nm. A spherical mirror focused the coronal light onto an annular fiber bundle, whose fibers were gathered and then illuminated the slit of an Ocean Optics

Spectrograph (see Figure 3). Simultaneous and co-aligned images of the corona were taken with a Sony video camera to follow the center of the field of view in the corona. The annular region over which light was spatially integrated was initially centered on the solar disk but also was repointed during totality to sample the outer F-corona at a distance of about 6 solar radii from disk-center.

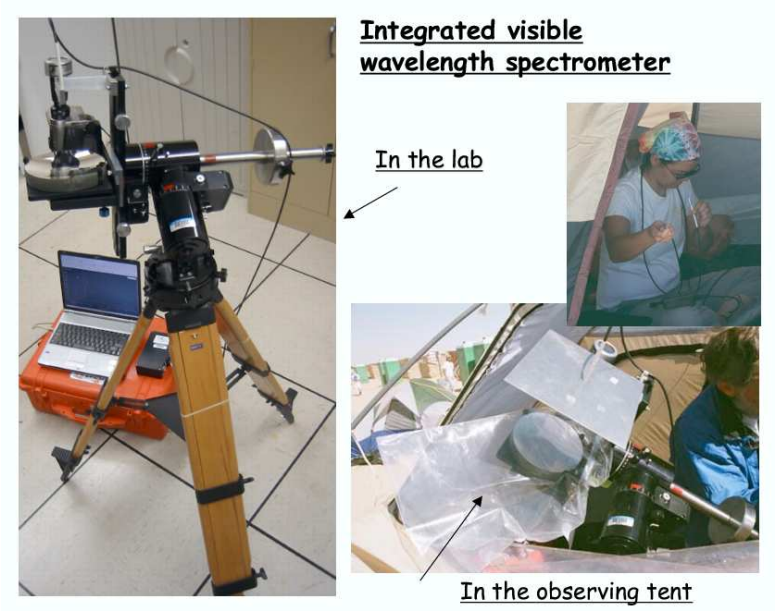


Fig. 3. The integrated visible wavelength spectrometer assembly in the lab (left), and in the eclipse observing tent (right).

2.3. The imaging polarimeter

Imaging of the corona was done simultaneously through three narrow (0.5nm) band-pass filters centered at 789.2, 1074.7 and 1083.0 nm, corresponding to the coronal emission lines of Fe XI, Fe XIII and He I respectively. In addition, a 50 nm wide filter, centered at 650 nm, was used for polarized brightness measurements of the white light coronal emission produced by electrons scattering photospheric radiation (see Figure 4). With the exception of the broadband filter, all filters were thermally controlled. Polarization measurements were made with all filters except 789.2. Each polarizer was rotated by 45° between exposures. Two identical electro-thermally cooled PIXIS 1024BR cameras, manufactured by Princeton Instruments, were used. They were both mounted on the same equatorial mount, each fitted with two filters that could be interchanged during the observations. The cameras have a 1024×1024 CCD array, with peak quantum efficiency of 87% at 800 nm, dropping down to 5% at 1100 nm. The polarizers were placed ahead of the filters, which were mounted on Nikon 300 mm focal length lenses. The observations were taken with 4 positions of the polarizer, separated by 45 degrees, at normal incidence, and with the filters tilted by 7° for observations in the nearby continuum. A number of exposure times,

Fig. 4. Two views of the imaging polarimeter from within the observing tent. Two PIXIS 1024BR cameras are mounted on the same German Equatorial mount, with filters attached to each camera. The filters could be interchanged during the observing run. The polarizers were manually controlled.

varying from a fraction of a second to 14 seconds, were used.

3. Preliminary Results

Shown in Figure 5 is an example from the IR spectrum taken with NIIS in the different fibers, hence at different spatial locations (i.e. latitude and radial distance) in the corona. The most dominant lines are those of Fe XIII 1074.7 nm and He I 1083 nm. The next dominant lines are the S IX emission at 1252.5 nm and Si X at 1430 nm. There is a new strong unidentified line around 1660 nm. In addition to the polarized spectrum, the other novel outcome of the NIIS measurements was a near-IR coronal image reconstructed from the polarized intensity collected by each fiber.

An example of a spectrum taken with the integrated visible wavelength spectrometer is shown in Figure 6. The spectrum was recorded when the fiber bundle was centered at the Sun, and is dominated by emission from the prominence off the east limb. A number of spectral lines are identified, and some are unknown. Note that the Fe XIII 1074.7 nm line is missing while the He I 1083 nm line is very prominent. This is a result of the reduced quantum efficiency of the detector around 1000 nm. However, the signal from the He I 1083 nm emission suggests that, despite the very low response of the detector at that wavelength, the intensity of the He I emission from the prominence was strong enough to be measurable, as opposed to the 1074.7 nm line that was not.

An example of the polarized brightness images taken in the broad band (pB) and at 1074.7 nm narrow bandpass filters are shown in Figure 7. Also shown is

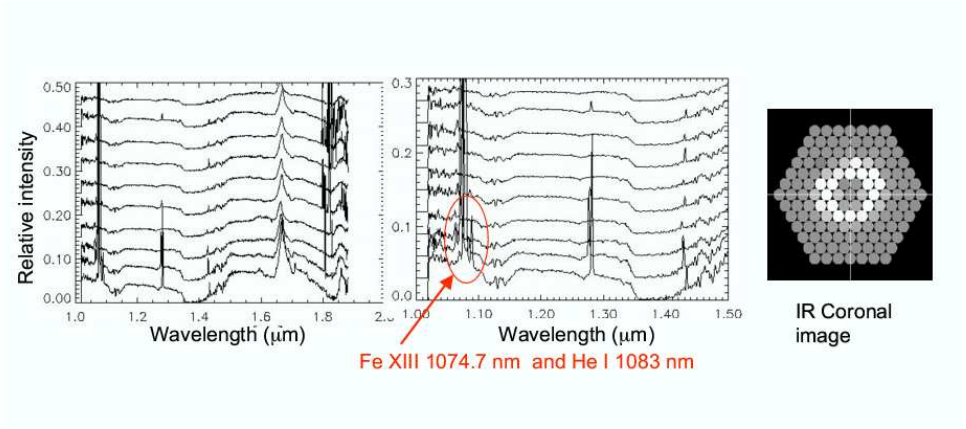


Fig. 5. Left: A sample near-IR spectrum covering 1-1.8 microns. The sample spectra collected from different fibers pointing at different regions of the corona, as shown in the IR coronal image to the right, are offset vertically in this plot. Middle: An enlargement of the left panel, with emphasis on the 1-1.5 micron wavelength range. The dominant emissions from the Fe XIII 1047.7 nm and He I 1083.0 nm lines are encircled. Right: A reconstructed image of the corona in the near IR from the integrated intensity in each fiber. Each circle represents the field of view of one fiber.

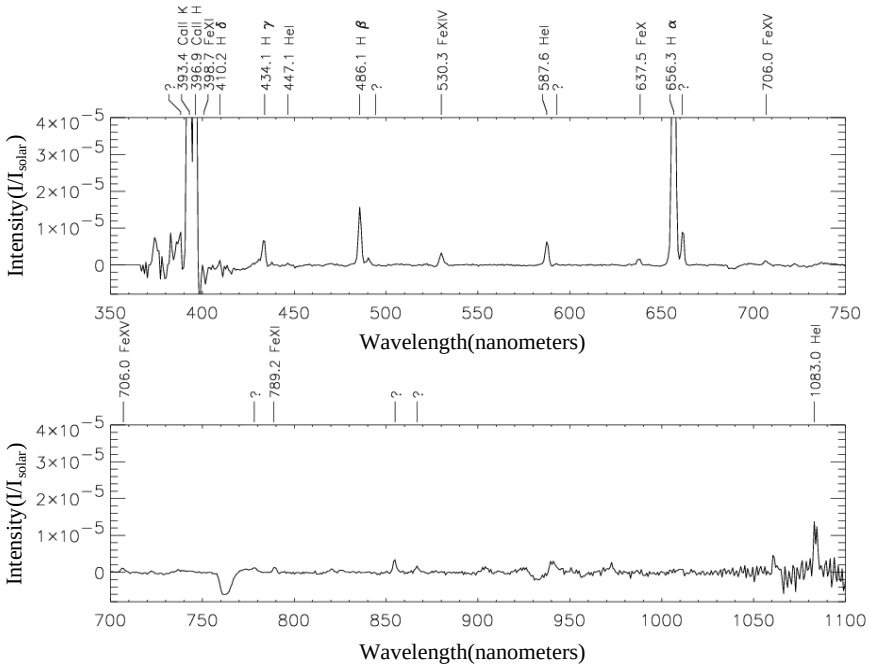


Fig. 6. An example of the *flattened* visible spectrum taken when the fiber bundle was centered at the Sun. The wavelength range covers 350–1100 nm. In this display the spectrum is divided equally in two parts with a repetition of the 700–750 nm range.

an image taken in Fe XI 789.2 nm which yielded a surprising result. The extent of the emission out to the edge of the field of view was quite unexpected, pointing to a dominance of the emission by resonant scattering beyond a fraction of a solar radius.

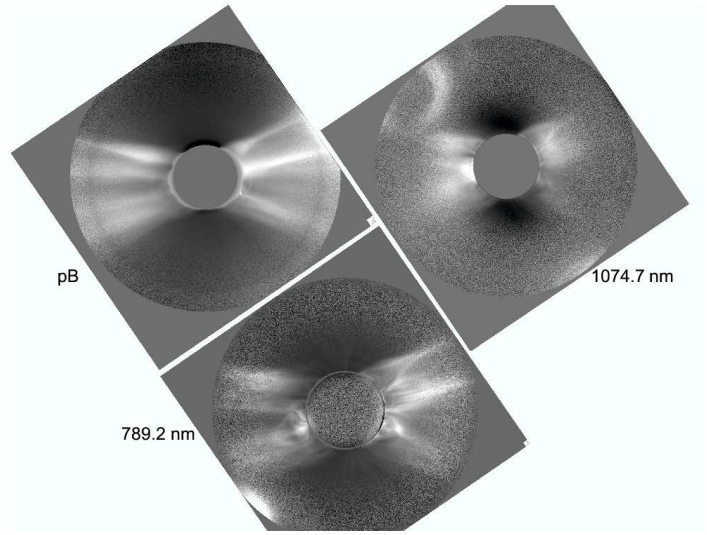


Fig. 7. Polarized brightness image (left) taken with the broadband filter centered at 650nm with a bandpass of 50 nm, polarized brightness image in 1074.7 nm (right) with a bandpass of 1 nm, and intensity image of the 789.2 nm Fe XI line (bottom panel). The orientation of the images is such that solar north is up.

4. Discussion and Conclusions

By all measures the eclipse observations made with the complement of three instruments were very successful. The results presented above are only preliminary, and data analysis is ongoing to answer the key questions that were the goal of the observations: (1) Can we detect and characterize a signature of nanometer size dust grains in the corona, (2) is there a neutral wind of interstellar origin in the corona, (3) can we characterize the near-Sun dust from F-coronal imaging and integrated visible spectroscopy, and (4) can we unambiguously identify the direction of the coronal magnetic field to place stringent constraints on theories of coronal expansion and solar wind acceleration? One totally unexpected result from these planned measurements, was the extended emission in the 789.2 nm Fe XI line, which has the potential to be a powerful diagnostic tool for future coronagraphic observations from both ground and space.

Acknowledgements. The members of the IfA eclipse group (see Figure 8) are grateful to Profs. Abdel Hamid Zeidan and Osama Shalabiea for their vision and relentless efforts that led to the

Fig. 8. Members of the IfA eclipse team at the observing site on 30 March 2006. Left to right, top row: Huw Morgan, Sarah Jaeggli, Judd Johnson, Shadia Habbal, Martina Arndt; bottom row: Adrian Daw, Jeff Kuhn, Ilia Roussev, and Don Mickey.

outstanding success of the eclipse observations at Waw AnNamous in Libya. We are also thankful to Abdel Razak Gherwash and Rob Schreuders for providing the necessary facilities at the eclipse camp site, in particular the liquid nitrogen and generators. We thank Alan Lichty from Princeton Instruments for enabling us to borrow a PIXIS near IR camera. We are also grateful to Diane Sakamato and Faye Uyehara for their help with instrument procurements and travel arrangements, to Lee Ann Scardina and Elizabeth Gutierrez, from DHL-USA, and Sayed Sabra and Mohammed AlSheybani, from DHL-Libya, for providing affordable and reliable shipping of the observing equipment to and from Libya, and to Appalachian State University Research Council for supporting the participation of A. Daw. Support for this project was provided by NSF grant ATM-0450799 to the Institute for Astronomy at the University of Hawaii.

References

Belomoin, G., *et al.*: 2002, *App. Phys. Lett.* **80**, 841
Charvin, P.: 1965, *Ann. d'Astrophys.* **28**, 877
Habbal, S. R., Arndt, M. B., Nayfeh, M. H., Arnaud, J., Johnson, J., Hegwer, S., Woo, R., Ene, A. and Habbal, F.: 2003, *ApJ*, **592**, L87
Kuhn, J. R., Jaeggli, S, Lin, H., and Arnaud, J.: 2006, COSPAR (in press)
Kuhn, J. R., Penn, M. J., and Mann, I.: 1996, *ApJ* **456**, L67
Mann, I., Krivov, A, and Kimura, H.: 2000, *Icarus* **146**, 568
Michels, J. G., *et al.* :2002, *ApJ* **568**, 385
Nayfeh, M. H. *et al.*: 2001, *Appl. Phys. Lett.* **78**, 1131
Nayfeh, M. H., Habbal, S. R., and Rao, S.: 2005, *ApJ* **621**, L121
Witt, A. N., Gordon, K., and Furton, D. G.: 1998, *ApJ* **501**, L111

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

EKPOL: LIQUID CRYSTAL POLARIMETER FOR ECLIPSE OBSERVATIONS OF THE K-CORONA

LUCA ZANGRILLI, SILVANO FINESCHI, GIUSEPPE MASSONE, GERARDO CAPOBIANCO and FRANCESCO PORCU
INAF - Osservatorio Astronomico di Torino, Strada Osservatorio 20, I-10025, Pino Torinese, Italy

and

PAOLO CALCIDESE
Osservatorio Astronomico della Regione Autonoma Valle d'Aosta, Nus, Saint-Barthélemy, Italy

Abstract.
The EKPol experiment is a K-corona imaging polarimeter, designed to measure the linear polarized radiation coming from the solar corona, during the total solar eclipse of March 29^{th}, 2006, in Waw an Namous - Libya. The EKPol polarimeter is based on Liquid Crystal Variable Retarders, allowing a polarimetric optics configuration without mechanically rotating retarders. This instrument is also thought as a proof of concept for a polarimeter to be placed along the visible light path of the UltraViolet and Visible-light Coronal Imager (UVCI) of the Sounding-rocket Coronagraphic Experiment (SCORE).

Key words: solar corona, K-corona

1. Introduction

The total solar eclipses offer a great opportunity to observe the faint solar corona, expecially its inner portions, which are not easily accessible by coronagraph telescopes, owing to the instrumental scattered light. As during eclipses, the corona is directly observed, those kind of observations benefit of a great reduction of instrumental scattered light, particularly in the inner solar corona close to the limb.

The continuum coronal emission in the visible light arises from Thomson scattering of the solar disc radiation by free electrons in the coronal medium (K-corona), and from the radiation scattered by coronal dust particles (F-corona). The brightness of the K-corona is directly proportional to the electron column density along the line of sight. By adopting proper geometrical assumptions (see e.g. van de Hulst, 1950), we can obtain information on the coronal medium electron density.

The K-corona emission turns out to be partially linearly polarized, as light scattered by coronal electrons is incident from a particular direction: unpolarized light is emitted isotropically from the solar disc and electrons scatter it. For a sufficient distant position of the scattering electrons, the Sun can be considered as a point source and an observer will see radiation with the electric field vibrations in a plane parallel to the solar limb.

In the following we describe the EKPol experiment, which is a K-corona telescope for the observation of the total solar eclipse of March 29^{th}, 2006, in Waw an Namous,

Libya. It consists of a Liquid Crystal Variable Retarder (LCVR) based polarimeter, and produces images of the linearly polarized radiation of the K-solar corona. The choice of using LCVRs allows replacing mechanically rotating retarders with electro-optical devices, without moving parts.

The experiment purposes can be summarized as follows:

- study the linearly polarized component of the K solar corona between 1.2 and 4 solar radii;
- observe the morphology of the coronal structures in the visible light;
- obtain the coronal electron density distribution, from the measured polarized brightness (pB);
- proof of concept and LCVRs test for a polarimeter due to be assembled in a sounding rocket instrument (SCORE; Launch: end 2006 - beginning 2007 (after STEREO launch).

The absence of mechanically rotating retarder, makes the EKPol concept suitable for being implemented in space-borne coronagraphs, such as the SCORE sounding rocket experiment, a set of two UltraViolet Coronagraph Imager (UVCI) with co-aligned and co-registered fields of view. UVCI is an externally occulted, off-axis Gregorian telescope, designed to provide full images of the extended corona and optimized for the HeII λ 30.4 nm and HI λ 121.6 nm narrow-band coronal emission (see e.g. Fineschi, 2003, Romoli, 2003).

EKPol has been designed and assembled in the Optics Laboratory of the Astronomical Observatory of Torino, in collaboration with the Osservatorio Astronomico della Regione Autonoma Valle d'Aosta.

2. EKPol description

The EKPol assembly essentially consists of:

- an achromatic doublet with 600 mm of focal length and 50 mm of aperture acting as objective lens;
- a LCVR based polarimeter, without moving retardance plates;
- a CCD camera detector, Pixel Vision, back illuminated, triple Peltier cooling;
- a Losmandy G11 telescope mount, with Gemini computer guide system.

The EKPol field of view is about 2°, and the solar corona can be detected up to 4 solar radii. The CCD camera pixel size is 24 μm, with a angular resolution of 8.6 arcsec per pixel. A design showing the EKpol mechanical structure is shown in Figure 1. As can deduced from Figure 1, the project of the structure is modular, allowing to test separately in the laboratory the collimator and camera lenses and the polarimetric optics complex. Moreover, the light baffles can be removed and replaced with other stops, defining the effective telescope aperture.

2.1. EKPol optical design

The construction specifics of EKPol are summarized in the table shown in Figure 2, and a scheme of the optical assembly of the EKPol instrument is shown in Figure 3. It essentially consists of off the shelf elements: the primary lens is an achromatic doublet from Melles-Griot; the collimating and camera triplets are from Optec; the polarimetric optics are from Meadowlark Optics, and the color filter is from Andover.

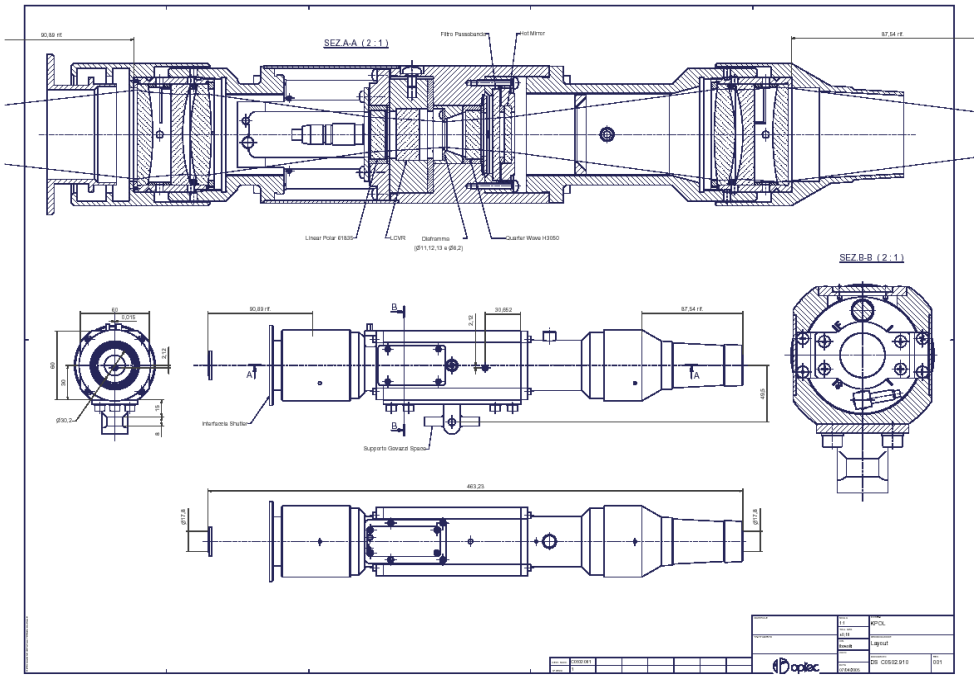


Fig. 1. EKPol mechanical structure; details of the optical mounts and polarimeter module are shown. In the configuration used during the eclipse campaign, the light baffle on the right has been properly replaced with another field stop.

The improvement of the optical design has been developed with the ZEMAX ray tracing program. The geometrical spot size, evaluated with ZEMAX is shown in Figure 4. Different spots for six fields (the remaining five fields are symmetric) are shown, corresponding to heliocentric distances of 0, 1.9, 2.1, 2.6, 3.2, and 4.4 solar radii. The fields relevant to the corona observations correspond to rms spot diameters between 5 and 13 μm, well within the 25 μm detector pixel size.

2.2. EKPol polarimeter design

The innovation of the EKPol polarimetric group consists of using a nematic Liquid Crystal Variable Retarder plate (LCVR), in a rotator configuration, allowing to replace mechanically rotating retarders with electro-optical devices without moving parts (for an analogous solar corona polarimeter design see also Elmore, 2000). Nematic liquid crystals are optically anisotropic media, acting locally as a uniaxial retarders and exhibiting optical birefringence. The effective birefringence can be changed by varying an applied voltage, producing different values of the retardance. The external applied variable voltage, ranging from 0 to 10 mV, with retardance values going from about λ to 0. The LCVR is constructed using precision polished, optically flat fused silica windows spaced a few microns apart, with the cavity filled with nematic liquid crystal material and sealed. The assembly ensures excellent transmitted wavefront quality and low beam deviation. The extraordinary (or slow)

E-Kpol Assembly

Polarization Analysis	Narrow Band Imaging of Linearly Polarized Visible-Light	
Optical Configuration	Doubly Telecentric Relay System (1:1 magnification)	
Imaging Optics (Eyepiece & Camera Lens)	Type: Pair of Achromatic Triplets (for each optics)	
	e.f.l.: 100 mm	
	Diameter: 50 mm	
	Achromatic Triplet	Optec design
		Glasses: N-LAF21/-SF10/N-LAK8
		Focal Length: 100 mm
Polarizing Optics	Achromatic λ/4 Retarder -MLO	Mod.: AQM-100-λ
		Diameter: 25.4 mm
		Clear aperture: 17.8 mm
	Liquid Crystal Variable Retarder (LCVR) - MLO	Mod.: LRC-200
		Diameter: 50.8 mm
		Clear aperture: 17.8 mm
		Retardance Range: 0 - (3/4) λ
	Linear Polarizer - MLO	Mod.: DPM-100 VIS2
		Diameter: 25.4 mm
		Clear aperture: 17.8 mm
Narrow band filter	Diameter: 40 mm	
	Bandpass : 580 – 660 nm (center 620nm)	
Objective lens	Achromatic Doublet	Glasses: N-BK7/N-SF5
		Focal Length: 600 mm
		Diameter: 50 mm

Fig. 2. EKPol construction specifics for the different instrumental modules.

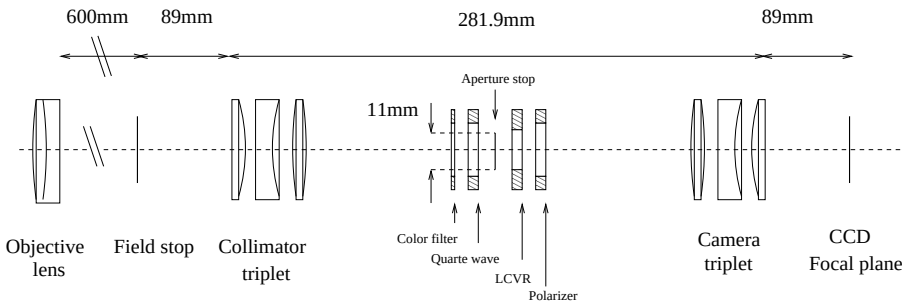


Fig. 3. EKPol optical diagram, showing the system sizes and the different groups of optics.

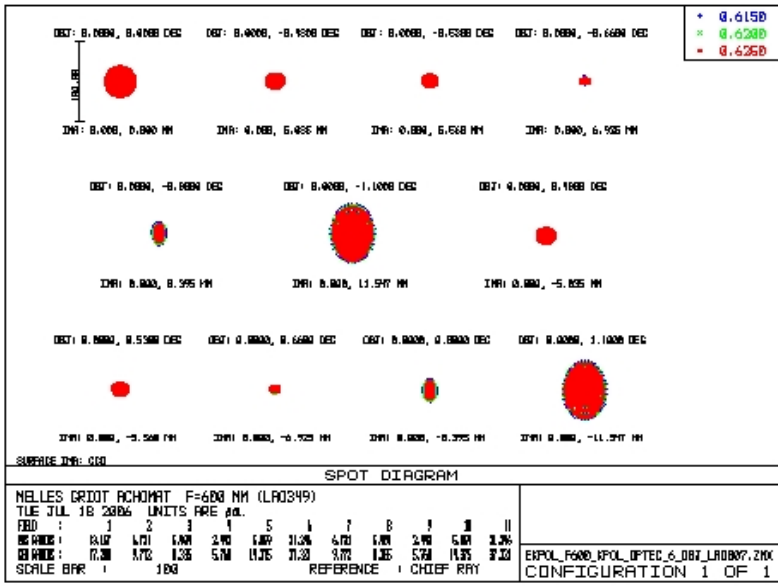


Fig. 4. Spot diagrams for different fields. The rms spot radii are reported.

index is defined by the long axis of the liquid crystal molecules. When no voltage is present, the molecules lie parallel to the windows and maximum retardance is obtained. When voltage is applied across the liquid crystal layer, the molecules tip parallel to the applied electric field. As voltage increases, the effective birefringence decreases, causing a reduction in retardance. These liquid crystal retarders are sensitive to temperature and wavelength changes, and stability of performances are obtained by a constant heating control and a narrow band selection.

The LC low-voltage electro-optical modulation of the polarization signal is accurate, reproducible, and - most important - fast, i.e., up to 100 Hz. Moreover, the control hardware is compact, light-weight, with limited power consumption. This is a clear advantage over more classical methods of mechanical modulation by rotation of polarizing elements, or of piezo-elastic modulation which require high driving voltages (i.e., kilovolts), while in the case of LCVRs large retardance intervals are obtained without requiring very high voltages. Using LCVRs, errors caused by moving parts, unavoidable when using mechanical rotation (inertia with its acceleration and brake times, misalignments of the signal on the detector, etc.), are eliminated. Precise rotation stages, stepper motors and accessories such as a gearbox are avoidable.

Owing to the chromaticity of LCVRs, for a given applied voltage different retardance values are produced within the visible band in corona (about $400 - 650$ nm), diminishing the contrast while modulating the polarized signal in a polarimeter. Hence, in principle, it would be necessary to characterize the LCVR chromatic response when observing non-monochromatic radiation. In EKPol, the bandpass is restricted by a narrow band filter, and the effect of integrating over the transmission

band results in a negligible loss of contrast. For a discussion about the chromaticity of LCVRs, see (Fineschi, 2003), where the results of first tests on the LCVR's chromatic response in a polarimeter assembly are presented.

The polarimeter concept is illustrated in Figure 5. It is constituted of a fixed

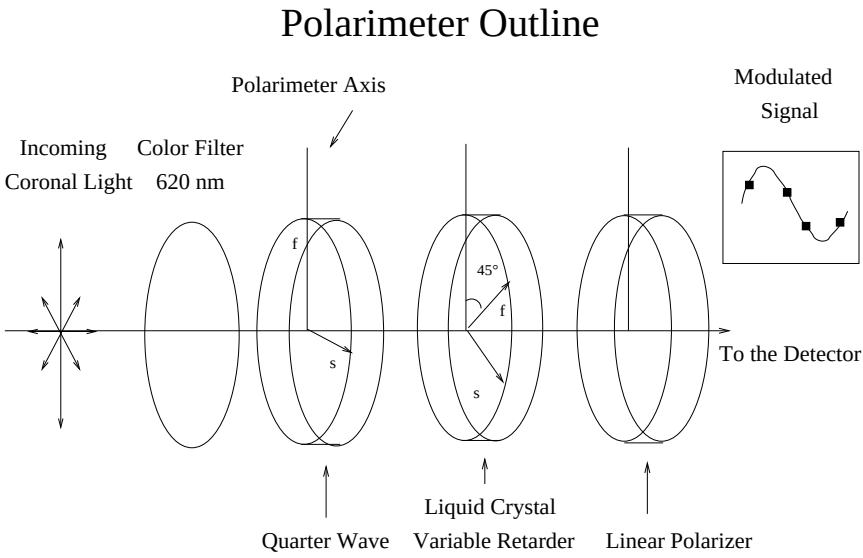


Fig. 5. Illustrating scheme of the EKPol polarimeter, with the essential optical components.

achromatic $\lambda/4$ retarder with the fast axis aligned with the 0° axis of the system, a LCVR with the fast axis at 45°, which allows the modulation of the polarized signal from the corona, and a linear polarizer with its transmission axis at 0°. This polarimetric configuration does not suffer from the possible presence of circularly polarized radiation. The polarimetric elements are placed in a telecentric beam, so transmission and polarization dishomogeneities are averaged across the field of view. The signal modulation, which is required to determine the Q and U Stokes coefficients, is obtained by varying the voltage applied to the LCVR, and taking exposures at four different retardance values. A narrow band color filter, centered about λ 620 nm, is placed in front of the polarimetric complex and selects the polarimeter bandpass, in order to minimize the pollution effect of the sky brightness.

In the configuration above described, the four measured signals m_i from the polarized corona are given by equation 1:

$$m_i = g(I + Q\cos\delta_i + U\sin\delta_i) + b \tag{1}$$

where I, Q, U are the input Stokes parameters, δ_i are the LCVR retardances, g is a constant factor containing the transmission of the polarimetric optics and the detector quantum efficiency, b is the detector bias and dark current. Linear combinations of the four measures for each pixel are used to obtain I, Q and U, and the polarized brightness is given by

$$pB = \sqrt{Q^2 + U^2} \tag{2}$$

Details about the instrument calibration procedure and data analysis will be given in a forthcoming paper, describing the results from this eclipse campaign.

2.3. Instrument control setup

The EKPol operation control has been made with a Compaq Evo workstation with Windows/NT operating system, for the CCD camera, and a Asus laptop with Windows/98, for the LC voltage and temperature settings. The LCVR output voltage was driven by a programmable digital controller, by LPT port, generating a 2 KHz frequency square wave. We acquired 16 bit, 1024x1024 pixel frames, from the CCD camera, connected by optical fibers from the digital converter to the PC board. The CCD camera properties are summarized in Table I:

TABLE I
CCD camera parameters.

Pixel size	24 μm
CCD op. temp.	237 K
Field of view	1024 x 1024 px
RO frequency	454 KHz
Dark Current	0.49 e^-/px/sec

3. Instrument characterization and observations

During the days preceding the eclipse, several tests on the control software and calibration operations where performed. The calibration procedures, included flat fielding, electronic bias and dark current measurements, absolute calibrations and polarimeter response. Focusing tests were made on star images and during the partiality phase of the eclipse on the lunar edge on the solar disc. Absolute calibration data has been taken by putting an opal in front of the telescope, on uneclipsed Sun at the same elevation as totality, taking exposures for each of three exposure times selected for the observations. A first choice of exposure times were made during preliminary working tests of EKPol, performed at the beginning of March at the Osservatorio Astronomico della Regione Autonoma Valle d'Aosta (Italy), on the full moon.

The polarization calibration setup included a prepolarizer plus an opal, and we took exposures at five positions of -90^o, -45^o, 0^o, 45^o, 90^o, and for each of the the

LCVR voltages of 10, 7, 5.4, and 4.5 V, which were used during coronal observations, and corresponding to an approximate LCVR retardance (after a first data analysis) of 66°, 144°, 242°, 323°, (referred to the polarimeter 0° axis).

Owing to keep the image registration error within the telescope resolution, the telescope mount was accurately stationed by using a standard Bigourdan procedure and a STV monitoring station with CCD camera. We got an error of 10 arcsec (about one detector pixel) during the 4 minutes of totality, on the night preceding the eclipse. Major difficulties where encountered with the ground solidity (the telescope mount tended to sink in the sand and suffered from vibration transmission), and the structure vibrations induced by the wind. However, during the totality phase the wind calmed down, and a first image inspection told us that the instrument maintained the coronal images in the required position.

The stability of the LCVR response has been ensured by keeping the crystal temperature constantly at about 30 °C, by means of a electric resistance heater.

The scheme of the observing sequence is given in the following Table II.

TABLE II
Observing sequence scheme adopted during eclipse observations; exposure times and LCVR voltages are given.

Exp (ms)	LCVR (V)	Seq.	Exp (ms)	LCVR (V)	Seq.	Exp (ms)	LCVR (V)	Seq.
250	10	first img.						
250	10	1	1000	10	2	4000	10	3
250	7	1	1000	7	2	4000	7	3
250	5.4	1	1000	5.4	2	4000	5.4	3
250	4.5	1	1000	4.5	2	4000	4.5	3

To get a further calibration and an idea of possible instrumental variations, polarization response, absolute calibration, bias and dark current measurements have been repeated immediately after the totality phase.

4. Conclusions

The EKPol experiment adopts the innovative technology of LCVRs, in a polarimeter assembly for the K-corona observation during the total solar eclipse of March 29th, 2006, in Waw an Namous - Libya. A first lookup to the data we obtained, confirms the validity of the instrument approach, which, in the next future, will be implemented in a polarimeter for the HERSCHEL/SCORE sounding rocket mission.

References

D. F. Elmore, et al.: 2000, in *Instrumentation for UV/EUV Astronomy and Solar Missions*, Proc. *SPIE*, **4139**, 370

S. Fineschi, et al.: 2003, in *Innovative Telescopes and Instrumentation for Solar Astrophysics*, Proc. *SPIE*, **4853**, 162

S. Fineschi, et al.: 2005, in *Solar Physics and Space Weather Instrumentation*, Proc. *SPIE*, **5901**, 389
M. Romoli, et al.: 2003, in *SOLAR WIND TEN; AIP Conference Proceedings*, **679**, 846
van de Hulst, H., C.: 1950, *Bulletin of the Astronomical Institute of the Netherlands*, **11**, 135


Panasonic


group

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

SEARCH FOR POSSIBLE SOLAR NEUTRINO RADIATIVE DECAYS DURING TOTAL SOLAR ECLIPSES

S. CECCHINI[*], D. CENTOMO, G. GIACOMELLI, R. GIACOMELLI and V. POPA[†‡]
Dipartimento di Fisica dell'Università and INFN Sezione di Bologna, I-40127 Bologna, Italy

Abstract. Total solar eclipses (TSEs) offer a good opportunity to look for photons produced in possible radiative decays of solar neutrinos. In this paper we briefly review the physics bases of such searches as well as the existing limits for the ν_2 and ν_3 proper lifetimes obtained by such experiments. We then report on the observations performed in occasion of the 29 March 2006 TSE, from Waw an Namos, Libya.

Key words: solar neutrinos, total solar eclipses

1. Introduction

In the last years, the evidence of neutrino (both solar and atmospheric) oscillations clearly shown that neutrinos have non-vanishing masses, and that the neutrino flavor eigenstates (ν_e, ν_μ and ν_τ) are superpositions of mass eigenstates (ν_1, ν_2 and ν_3) (Ambrosio M.*et al.*, 1998; Fukuda Y. *et al.*, 1998; Ahmad Q.R. *et al.*, 2001). Neutrinos could undergo radiative decays, e.g. $\nu_2 \to \nu_1 + \gamma$, as initially suggested in (Melott and Sciama, 1981); the present status of the neutrino decay theory is summarized in (Sciama, 1995).

The neutrino radiative decay requires a non-vanishing neutrino magnetic moment; very stringent existing limits ($\mu_\nu < 1.3 \times 10^{-10} \mu_B$, (Hagiwara *et al.*, 2002)) refer to neutrino flavor eigenstates and thus are not directly applicable to possible dipole magnetic moments of neutrino mass eigenstates.

Neutrino decays (radiative or not) were searched for from astrophysical phenomena such as Supernova physics or the absence of γ rays in the Sun radiation, or from infra-red background measurements. Such neutrino lifetime lower limits are typically very large (e.g $\tau_0/m > 2.8 \times 10^{15}$ s/eV where τ_0 is the lower proper lifetime limit for a neutrino of mass m, (Bludman, 1992)), but they are indirect and rather speculative.

Much lower "semi-indirect" limits were obtained from the re-interpretation of solar and atmospheric neutrino data. Although the present accepted interpretation of the existing observations is essentially that of neutrino oscillations, the hypothesis of neutrino decays cannot be completely discarded, as a secondary effect. As an example, from the SNO solar neutrino data, in (Bandyopadhayay *et al.*, 2003) a

[*] Also IASF/CNR, I-40129 Bologna, Italy
[†] Also Institute for Space Sciences, R-77125 Bucharest Măgurele, Romania
[‡] presenting author

lower limit of $\tau_0/m > 8.7 \times 10^{-5}$ s/eV was inferred. By combining all available solar neutrino data, limits as $\tau_0/m > 2.27 \times 10^{-5}$ s/eV (Joshipura *et al.*, 2002), or, following different assumptions, $\tau_0/m > 10 \times 10^{-4}$ s/eV (Beacon *et al.*, 2002) were obtained.

Direct searches for radiative neutrino decays (not correlated with TSEs) were also performed: in the vicinity of nuclear reactors (e.g. (Bouchez *et al.*, 1988), yielding limits between $\tau_0/m > 10^{-8}$ s/eV and $\tau_0/m > 0.1$ s/eV, for $\Delta m/m$ between 10^{-7} and 0.1), or using the Borexino Counting Test Facility at Gran Sasso, at the level of $\tau_0/m \simeq 10^3$ s/eV, as a function of the neutrino polarization (Derbin A.V. *et al.*, 2002).

The Sun is a very strong source of ν_e neutrinos; the expected flux at the Earth (neglecting oscillation effects) is $\Phi \simeq 7 \times 10^{10}$ cm^{-2}s^{-1}. In normal conditions, if radiative neutrino decays occur yielding visible photons, they would not be observable due to the very large amount of light produced by the Sun. During a TSE, the Moon absorbs the direct light from the Sun, but is completely transparent to solar neutrinos. An experiment looking for such an effect is thus sensitive to neutrino decays occurring in the space between the Moon and the Earth.

In a pioneering experiment performed in occasion of the October 24, 1995 TSE, a first search was made for visible photons emitted through radiative decays of solar neutrinos during their flight between the Moon and the Earth (Birnbaum *et al.*, 1997). The authors assumed that all neutrinos have masses of the order of few eV, $\delta m^2 \simeq 10^{-5}$ eV^2, and an average energy of 860 keV. Furthermore, they assumed that all decays would lead to visible photons, which would travel nearly in the same direction as the parent neutrinos. In the absence of a positive signal, the search yielded a lower proper lifetime of 97 s.

In (Cecchini *et al.*, 2004a) we discussed a complete Monte Carlo model for the solar neutrino radiative decay. The neutrino production inside the core of the Sun is described according to the Standard Solar Model (SSM) predictions (Bahcall, 1998), both from the point of view of the energy spectrum and of the source geometry. We considered the ν_1 mass m_1 between 10^{-3} and 0.3 eV, and ν_1 the lowest mass neutrino state. The mass differences $\delta m^2 = m_2^2 - m_1^2 = 6 \times 10^{-5}$ eV^2 and $\Delta m^2 = m_3^2 - m_2^2 \simeq m_3^2 - m_1^2 = 2.5 \times 10^{-3}$ eV^2 are chosen in agreement with the solar and atmospheric neutrino oscillation experimental data, respectively. The mixing angles were also considered as measured by oscillation experiments, and the still unknown θ_{13} was chosen as $\sin^2\theta_{13} \simeq 0.1$. In calculating the zenith angular distributions of the emitted photons we considered the probability density functions (pdfs) (corresponding to 3 different neutrino polarizations: $\alpha = -1$ (lefthanded neutrinos), $\alpha = 0$ (Majorana neutrinos) and $\alpha = 1$ (righthanded neutrinos). The pdfs were integrated according to the angular resolution of the simulated experiments. The simulations yielded narrow visible signals corresponding to the solar $\nu_2 \rightarrow \nu_1 + \gamma$ decays, concentrated in about 100 arcsec in the direction of the center of the Sun. The signal corresponding to $\nu_3 \rightarrow \nu_{1,2} + \gamma$ decays is expected to be broader, with annular maxima at about 250 arcsec away from the direction pointing to the center of the Sun.

We made our first search for solar neutrino radiative decays during the June 21, 2002 TSE, in Zambia (Cecchini *et al.*, 2004b). Two data sets were obtained

and analyzed: 4149 frames recorded by a digital video-camera equipped with a 2× telelens and a 10× optical zoom, and 10 digital photographs made with a Maksutov-Cassegrain telescope (90 mm Φ and $f = 1250$ mm). The proper lower time limits (95% C.L.) obtained for the $\nu_2 \to \nu_1 + \gamma$ decays of lefthanded neutrinos range from $\tau_0/m_2 \simeq 10$ to $\simeq 10^9$ s/eV, for 10^{-3} eV $< m_1 <$ 0.1 eV. These limits are among the best obtained from direct measurements, demonstrating the potentiality of neutrino decay experiments during TSEs.

In this paper we report on the observations performed in occasion of the 29 March 2006 TSE, from Waw an Namos, Libya. We describe the experimental set-up and make a first characterization of the obtained data. The data analysis in terms of solar neutrino radiative decays is an ongoing lengthy process, so we cannot give here estimates on the lifetimes or lower limits.

2. 29 March 2006 total solar eclipse observation

Considering the good quality of transport and on-site facilities provided by the organizers of the Eclipse event in Libya, we prepared a main experiment designed to reach a sensitivity improvement of at least three orders of magnitude relative to our 2002 observations (Cecchini *et al.*, 2004b). For redundancy, smaller back-up experiments were also prepared.

2.1. The main experiment

Our main experiment used a Maksutov-Cassegrain telescope ($\Phi = 235$ mm, $f = 2350$ mm), equipped with a fast Mx916 CCD camera. The original findscope was substituted by the same digital videocamera that we used for data taking during the 2002 TSE. Figure 1 shows our apparatus, mounted at the observation site; the solar filter was removed during the totality phase.

The night before the TSE we aligned the system, adjusted the focus and took calibration images of some standard luminosity stars. In order to avoid the over-heating of the telescope and CCD during the eclipse day and to minimize the possibility of focus and alignment changes, the equipment was protected with aluminum foils.

The telescope movement was set to follow the Sun, in order to have always the center of the acquired images coincident with the Sun center. Furthermore, we implemented a special CCD exposure algorithm, in order to adapt the exposition times to the luminosity level of the Moon image. Although the ashen light is one of the main background sources in the search for a signal produced by solar neutrino radiative decays, it allows the reconstruction, frame by frame, of the real position of the Sun behind the Moon, eliminating the risk of pointing errors due to undesired movements of the telescope tripod on the sand. The algorithm analyzed in real time the luminosity of the previous registered frames, and determined the exposure time for the next frame for an average luminosity of the image at half value of the CCD dynamical range, and maximizing the contrast.

In Fig. 2 we present a comparison between a full Moon image (left panel) and two frames registered by our experiment at the beginning of the totality phase (middle panel) and near to its end (right panel). The approximate field of view of our observations is also marked by the rectangle on the full Moon picture.

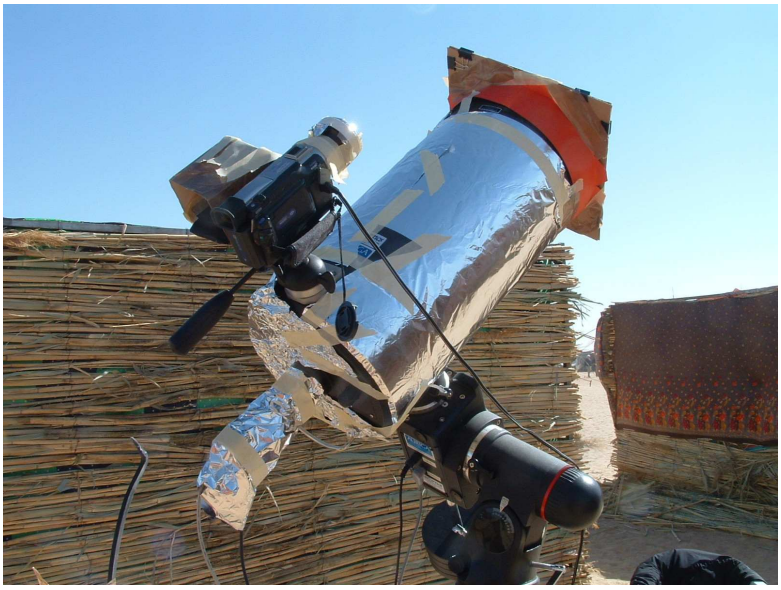


Fig. 1. The main experiment installed on the observation site, in the Waw an Namos eclipse camp. See text for details.

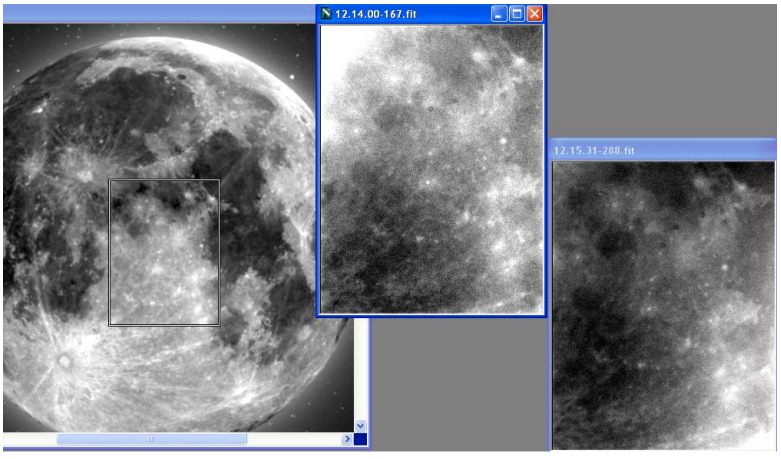


Fig. 2. Two frames obtained by our experiment at the beginning of the totality phase (middle) and near to its end (right), compared with an image of the full Moon (left). The approximate field of view of our observations is also shown on the full Moon picture

The luminosity change between frames may come from diffraction of the coronal light in the Earth's atmosphere. The effect was taken into consideration by our automatic exposure algorithm, as shown in Fig. 3.

For most of the frames, the exposure time was about 0.5 s; some attempts were made by the algorithm to increase the exposure time to about 1 s, all of them after

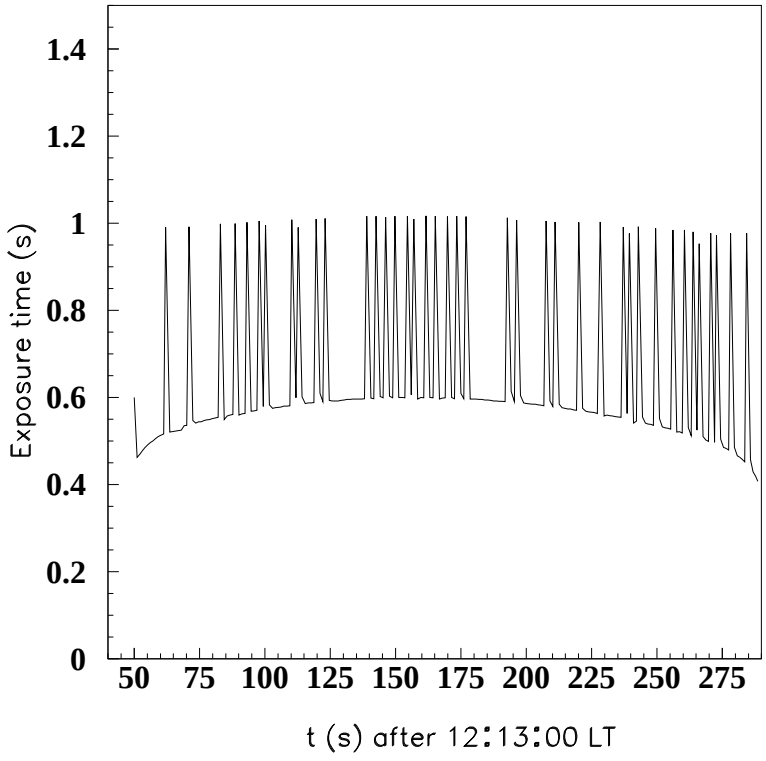


Fig. 3. The distribution of the CCD exposure time, during the totality phase of the eclipse.

a frame that had a lower luminosity than the required value. This could also be due to some fluctuations in the electric power affecting the acquisition, but since we have all the information concerning the particular conditions in which each frame was recorded, this will not interfere on further analyses.

In all panels of Fig. 2 a number of small luminous craters is easily observable. We selected 7 of them, to be used as "fiducial points" in order to test the telescope movements and to reconstruct the position of the Sun for each of the about 200 frames taken. The displacement of those markers during the totality phase is shown in Fig. 4; note that the North direction is from left to right, due to the CCD relative position.

From Fig. 4 it is clear that the fiducial points represent the Moon's movement in front of the eclipsed Sun, with some deviations in some frames taken in the middle and at the end of the totality. Those small pointing jumps were probably due to human activity in the vicinity of the telescope, but they may be corrected using the actual coordinates of the reference craters.

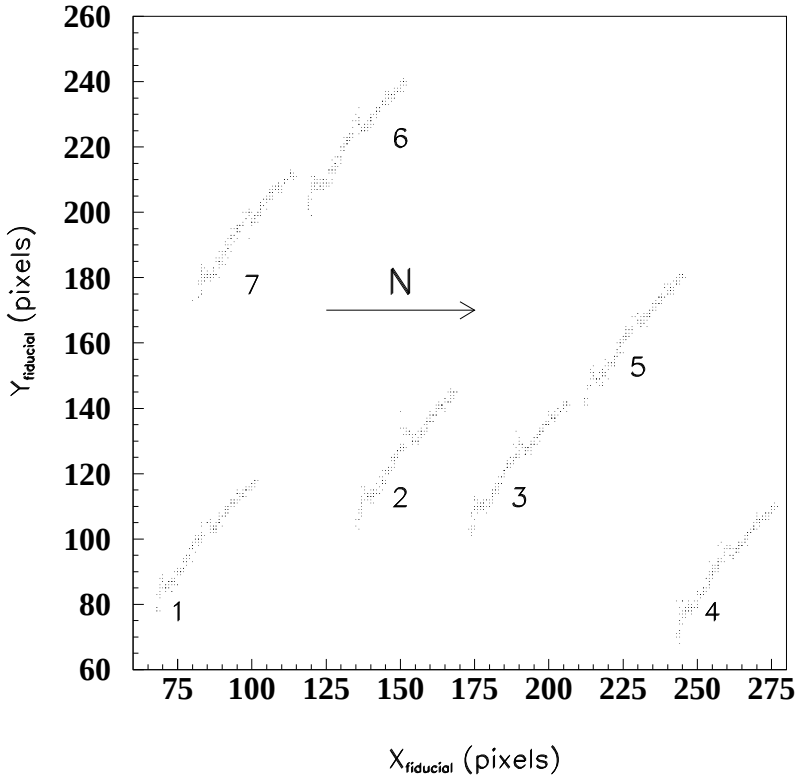


Fig. 4. The movement of the selected "fiducial points" (small luminous craters on the Moon surface) during the totality phase of the eclipse. The arrow indicates the North direction.

2.2. The backup experiments

The digital video-camera (the same as that used in occasion of the 2002 TSE) was itself a small backup experiment. It produced a digital film of the TSE, that could at least confirm our earlier results (Cecchini *et al.*, 2004b).

We also used a smaller Celestron C5 telescope, equipped with a manually controlled digital camera (Canon D20), as shown in Fig. 5. We obtained 50 digital pictures of the eclipse, as the one presented in Fig. 6; they will be analyzed in the same way as reported in (Cecchini *et al.*, 2004b). All those pictures are available from the science popularization site http://www.scienzagiovane.unibo.it.

3. Conclusions

In occasion of the March 29 2006 total solar eclipse we made several observations from the Waw an Namos camp, in southern Libya, looking for a visible light sig-

Fig. 5. The backup telescope with the digital camera, on the observation site.

nal possibly produced in solar neutrino radiative decays, during the neutrino flight between the Moon and the Earth.

Our main data consist of a set of about 200 CCD images of the central area of the Moon, aligned with respect to the position of the center of the Sun. The performant instrumentation used and the specially developed algorithm optimizing the exposure time allowed to visualize on each frame clearly recognizable details of the Moon surface, in the light reflected by the Earth. This will allow a precise reconstruction of the relative position of the Sun and Moon in each frame, thus eliminating the pointing errors due by accidental movements in the vicinity of the apparatus. Furthermore, as the ashen light is one of the major background sources in our search, the quality of the obtained images suggests that we can reach an experimental sensitivity close to the best possible in measurements using the visible spectrum.

The backup experiments worked also very well: we have 50 digital photographs of the total solar eclipse, obtained with a smaller telescope and digital films obtained with two digital videorecorders, that may also be used for pedagogical purposes.

The main data are currently under analysis; if no positive signal will be found,

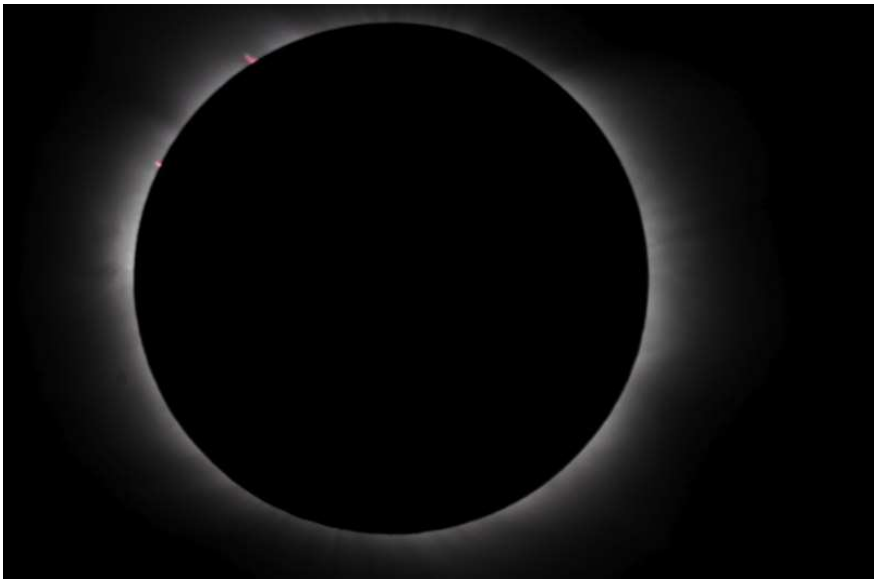


Fig. 6. One of the 50 digital photographs obtained with the backup experiment.

we expect a major improvement of our previous limits (Cecchini *et al.*, 2004b). They will also allow a precise measurement of the Earth's albedo.

Acknowledgments. We are grateful to the organizers of the SPSE 2006 event, for all their efforts that allowed to perform our experiment and to join the very exciting Symposium in Waw an Namos. Special thanks are due to the Winzrik Group and to the Libyan Air Force, for their kind and efficient assistance.

References

Ahmad Q.R. *et al.* (SNO Coll.): 2001, *Phys. Rev. Lett.* **87**, 071301
Ambrosio M. *et al.* (MACRO Coll.): 1998, *Phys. Lett.* **B 434**, 451
Bahcall J.N.: 1998, astro-ph/9808162
Bandyopadhayay A., Choubey E. and Goswami S.: 2003, *Phys. Lett.* **B 555**, 33
Beacon J.F. and Bell N.B.: 2002, *Phys. Rev.* **D 65**, 113009
Birnbaum C., Désauté, Foussard P., Nguyen M.G., Schwemling P. and Vannucci F.: 1997, *Phys. Lett.* **B 397**, 143
Bludman S.A.: 1992, *Phys. Rev.* **D 45**, 4720
Bouchez B. *et al.*: 1988, *Phys. Lett.* **B 207**, 217
Cecchini S., Centomo D., Giacomelli G., Popa V. and Şerbănuţ C.G.: 2004a, *Astropart. Phys.* **21**, 35
Cecchini S., Centomo D., Giacomelli G., Giacomelli R., Popa V., Şerbănuţ C.G. and Serra R.: 2004b, *Astropart. Phys.* **21**, 183
Derbin A.V. and Smirnov O.Ju.: 2002, *JETP Lett.* **76**, 483
Fukuda Y. *et al.* (Super-Kamiokande Coll.): 1998, *Phys. Rev. Lett.* **81**, 1562
Hagiwara K. *et al* (P.D.G.): 2002, Rev. **D 66**, 0100001
Joshipura A.S., Massó S. and Mohanty S.: 2002, *Phys. Rev.* **D 66**, 13008
Melott A.L. and Sciama D.W.: 1981, *Phys. Rev. Lett.* **46**, 1369
Sciama D.W.: 1995 *Nucl. Phys. Proc. Suppl.* **38**, 320

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

IS IT WORTH RE-DOING THE EDDINGTON EXPERIMENT? - A FEASIBILITY ASSESSMENT AT SOLAR ECLIPSES 2005-2006 IN LIBYA

JEAN-LUC L. J. DIGHAYE
EurAstro Association, Helfendorfer Str. 7/4, D-81671 Munich, Germany

Abstract. We describe an experimental set-up, comprising a full-frame digital single-lens reflex camera and an apochromat refractor, for measuring light bending using stars at the time of the annular solar eclipse of 2005 and the total solar eclipse of 2006 in Libya. We identify data reduction problems, and we propose further experiments. We ask for volunteers who could assist us in reducing the data.

Key words: Eddington, general relativity, solar eclipse

1. Introduction

In 1919, Sir Arthur Eddington attempted to measure the bending of the light emanating from stars situated in the background of the totally eclipsed Sun (Eddington, 1920), as foreseen by Einstein's newly discovered theory of general relativity. Results were somewhat controversial (von Mettenheim, 1996); however, they generally supported Einstein's theory. The bending effect, though, is tiny - it amounts to 1.75 arcsec at the solar limb, and it decreases as the inverse of the apparent distance of the star to the Sun's centre - and Eddington's set-up suffered from several error sources, one of them being the lack of temporal stability of the photographic plates used back them. The errors could only be mitigated by choosing a long focal length telescope, which in turn was rendered possible by the presence of many bright stars, belonging to the Hyades cluster, in the vicinity of the eclipsed Sun.

There have been few successful attempts to repeat Eddington's experiment since then. One example was the expedition to the eclipse of 30 June 1973 by the Texas Mauritanian Eclipse Team (Brune *et al.*, 1976; Jones, 1976). Heavy dedicated equipment was mandatory. Now, with large electronic detectors becoming available to amateur astrophotographers, it would be interesting to assess whether the inherent better stability of those detectors, together with the ease of data processing, would yield meaningful results of star displacements if the experiment would be re-done using telescopes of modest focal length. This should be considered as a technical achievement rather than further evidence towards Einstein's theory, now widely accepted, and for which much more accurate tests exist.

2. Location

In order to compare the disturbed positions of the stars (i.e. during the eclipse) with the undisturbed ones, the reference star field should be imaged during the

night, under conditions as close as possible to those of the eclipse, comprising:
- same height and orientation in the star field in the sky;
- same temperature and atmospheric pressure.

The first condition implies that the reference images should be taken from the same place as the eclipse field (or from an equivalent place of comparable latitude and elevation); a time separation of approximately 6 months between reference images and total eclipse is preferred, so that the reference field can be imaged in the middle of the night.

There is a fortuitous opportunity in Libya, occurring at the Annular Solar Eclipse (ASE) of 03 October 2005, and the Total Solar Eclipse (TSE) of 29 March 2006. Such eclipses occur at an interval of approximately 6 months, and the respective paths of the ASE and the TSE cross each other in a rather sunny, observation-friendly region of hills and cliffs called Djebel Maaruf. The coordinates of the intersection point are 25.156°N 18.576°E.

Strictly speaking, there is no need for an ASE six months before or after a TSE - reference images do not necessitate an eclipse. However, the present circumstances allow for a preliminary test of the TSE set-up during the ASE.

Libya, at the beginning of October, is warmer than at the end of March, but it is also warmer around the end of the morning - even with an eclipse occurring - than in the late evening. Thus, with some luck, the temperature would not be too different when reference images and eclipse images are taken. However, a calibration in temperature of the set-up is foreseen.

Whereas tour operators could freely organise expeditions to the Libyan desert at the time of the ASE, we were informed that only a few authorised operators would be allowed into restricted places at the time of the TSE. Anticipating such restrictions, the ASE expedition was complemented by a site evaluation in the area of the planned Eclipse-City camp (24.496°N 17.960°E) where the Solar Physics / Solar Eclipse (SPSE) international symposium was intended to be held.

3. Set-up (ASE'05) and error sources

Astrometric reference images were taken both at Eclipse-City and Djebel Maaruf, using:
- a Tele Vue 101/540 four-lens apochromat refractor, hereafter TV;
- a Canon EOS 1Ds Mark II digital single-lens reflex camera (size 24x36 mm, 16.7 megapixels), used at full resolution, raw+jpeg mode, ISO 100, noise reduction on, several images taken at 1/4s, 1s, 4s and 15s exposure times; and
- a Vixen equatorial mount.

Image registrations took place during the nights before 03 October 2005, at the time when the reference field had the same altitude and azimuth as during the upcoming TSE. Due to a power supply shortage, flat fields and dark exposures were taken afterwards, at the approximate midnight temperature in the desert i.e. 25°C, with the camera turned on about 1/4 hour in advance to simulate actual imaging conditions.

The reference star field was unfortunately poor in bright stars, so that a short focal length was used in order to include enough reference stars (see Figure 1), such

as:

– SAO 109297 (7.64 mag) & 109296 (8.02 mag) to the upper left
– SAO 109315 (6.40 mag) & 109278 (7.48 mag; close to the Sun) to the lower left;
– SAO 109206 aka CF Psc (ca. 6.89 mag) to the upper right;
– SAO 109216 (7.17 mag) to the lower right.

(Note: object NGC 128 is assumed to be invisible during TSE'06).

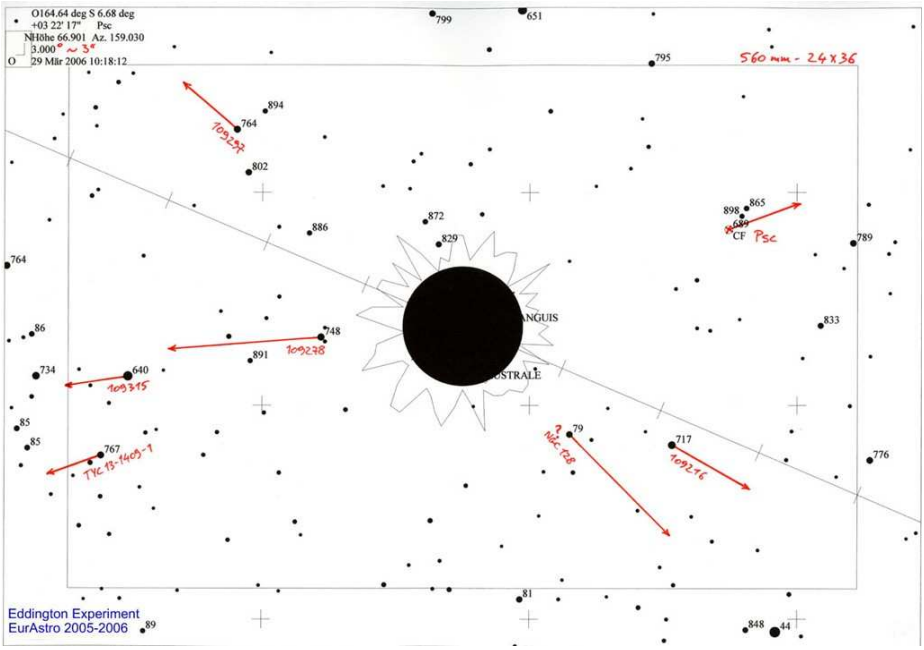


Fig. 1. Reference field showing exaggerated star displacements.

In Figure 1, the separation between vertical cross-hairs is 1 degree, whereas the arrows showing the displacement of some stars have the same scale for 1 arcsec, i.e. they are exaggerated 3600 times. All the displacements are below 1 arcsec, to be compared with the scale of the images which is 2.75 arcsec/pixel. Hence the centroid of the stars must be determined with sub-pixel accuracy, which requires a careful measurement error analysis, including, in addition to the factors envisaged by Jones (1976):

noise reduction: dark fields and flat fields must be acquired; the Canon sensitivity has to be set to ISO 100 which provides the best signal/noise ratio;

correct exposure: bracketing, especially during TSE'06, must be performed so that the pixel counts are sufficient, without saturating the detector. According to Espenak and Anderson (2004), correct exposure for the solar corona between 0.5 to 4 solar radii are comprised between 1/30 and 1/2 s at ISO 100;

distortion control: reference images were taken at several slightly different locations in the sky, so that the field closest to the actual images taken during

TSE'06 could be used; the other images were used too, in order to evaluate the incidence of optical distortion when reference images and eclipse images do not exactly coincide, and the systematic effect of atmospheric refraction;

temperature-induced scale changes: images of further reference star fields, taken at higher and lower temperatures, respectively, could be compared in order to assess the thermal expansion of the optical system (telescope, camera, detector). This need not be done in Libya - any star field being visible at the same place of the sky when such respective images are taken, e.g. during a warm summer evening and a cold winter dawn, will suffice;

atmospheric turbulence control: reference images should be compared and/or averaged to estimate the effect of atmospheric turbulence, or any other phenomenon related to anomalous atmospheric refraction, on astrometric positions (Hamammu, 2005). In view of the potential importance of that effect, a large number of reference images were taken in 2005.

4. Set-up (TSE'06) and preliminary assessment

The EurAstro expedition to Libya 2006 was carried out within the frame of the SPSE. The observation point was close to the nominal Eclipse-City location, and substantially the same set-up as for ASE'05 (see Section 3 above) was used. The sky was cloud-free, with little suspended sand and aerosol content. Temperatures varied between 21°C and 27°C during eclipse, therefore staying close to the thermal conditions of the reference images. Actual exposure times have been stored in the headers of all the images taken. Figure 2 is a processed, low-resolution composite image showing earthshine, solar prominences and coronal structure. Unprocessed full resolution single frames show stars down to about 8 mag, and processed composite images could yield even more stars usable for astrometry.

5. Data reduction strategy

Before TSE'06, the following strategy was considered:

- The visibility of the stars in the TSE images must be assessed first:

  A. if stars are visible in the TSE images, then proceed with data reduction;

  B1. if stars are not visible and the weather was the culprit, then try again at TSE'08;

  B2. if stars are not visible and the set-up is the culprit, then another set-up should be used for further experiments.

- Case A implies to evaluate the luminosity gradients in the images, and if possible to develop a mathematical model to remove such gradients. Then, in the correctly exposed images, the centroids of the stars must be measured as previously indicated, and their position corrected in view of the error sources previously identified. If there are residuals between the reference images and the eclipse images, and those residuals show a fair agreement with Einstein's predictions, then the experiment will be considered a success.

Fig. 2. Processed composite image taken during TSE'06.

– Cases B1 and B2 necessitate further experimentation. In particular, it is noted that the sky region next to the eclipsed Sun during the TSE of 01 August 2008 is richer in stars, including star cluster M44 (too far away, though, to be used for accurate measurements).

EurAstro will send at least one team for observing TSE'08, maybe from Mongolia or China. Since we do not intend to travel twice to Mongolia - or another remote place - with heavy equipment just for the sake of validating an experiment of little scientific importance, the reference star field should be imaged, about 6 months before or after TSE'08, from an appropriate European location.

A set-up for case B2 was already tested during a EurAstro mission to Crete in 2004. It comprises a LORVAF 200/1600 three-lens apochromat refractor and a M100B equatorial mount.

In view of the encouraging results obtained as stated in section 4 above, the data reduction can begin as foreseen in step A. This calls for coordinated activities.

6. Coordinated activities

The reduction of the data obtained before ASE'05 and at TSE'06 involves a lot of work, especially in view of the large number of usable images. Such work could be of rewarding didactic value to students or academic institutions. The author would be glad to send them, for instance by CD or DVD, all of our raw images. EurAstro

is an informal nonprofit association, therefore we intend to provide the information, and further support if needed, free of charge; also, most of our members are amateur astronomers, not professionals. If the results of the data reduction are of personal interest to the student or professional having performed them, we agree that he or she could use them freely.

Acknowledgements. The persons who made this project possible are acknowledged in the following link: http://www.eurastro.de/missions/sofi06/libya.htm

References

Brune Jr., R. A. *et al.*: 1976, AJ, **81**, 452
Eddington, A. S.: 1920, Space, Time and Gravitation - An Outline of the General Relativity Theory, Cambridge University Press, ISBN 0521337097
Jones, B. F. 1976, AJ, **81**, 455
von Mettenheim, C.: 1996, The Critical Rationalist **1**, No. 3, ISSN 1393-3809
Espenak, F., Anderson J.: 2004, NASA/TP-2004-212762
Hamammu, I. M.: 2005, personal communication

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

THE SUN AND SOLAR ECLIPSES IN THE CLASSROOM

MARTINA B. ARNDT
Department of Physics, Bridgewater State College, Bridgewater, Massachusetts, 02325, USA

Abstract. Educators in the United States are working hard to meet the challenges of the *No Child Left Behind* initiative signed by George Bush in 2002. The ultimate goal of this initiative is to improve the performance of America's elementary and secondary schools. In Massachusetts this has, in part, resulted in standardized testing (known as the MCAS exam) at various times throughout a child's schooling and a drive to provide professional development opportunities for teachers. In the physics department at Bridgewater State College (in Bridgewater, Massachusetts), we offer graduate level courses that count toward professional development for K-12 teachers. My goal is to develop a graduate course centered on the Sun that not only educates teachers, but also provides them with curricular material that prepares their own students for the MCAS exam.

Key words: solar physics education, eclipses

1. Introduction

In 2002, President G.W. Bush signed the *No Child Left Behind (NCLB)* initiative into law (www.ed.gov/nclb). The ultimate goal of this initiative is to improve the performance of America's elementary and secondary schools and close the achievement gap between students. The four main principles of this initiative are:

1. *Stronger accountability for results*: states must have a plan for closing gaps in student achievement and have ways to assess their plan.
2. *Freedom for states to use federal funds*: states can use federal funding to address particular needs, helping them reach their goals of closing the student achievement gap.
3. *Using proven education methods*: schools receive federal funding to employ good teaching practices that are based on research studies.
4. *Choices for parents*: parents have access to more information that helps them make informed decisions about their child's education and even have options to move their child to other schools.

As a result of this federal directive, each state has the flexibility to deal with disparities in student achievement in their own way. Massachusetts has chosen to address the first pillar of accountability by implementing the Massachusetts Comprehensive Assessment System (MCAS) (www.doe.mass.edu/mcas) and requiring that teachers be Highly Qualified (www.doe.mass.edu/nclb/hq).

1.1. The MCAS exam

The MCAS is an exam administered to students at various points in their studies between the 3rd and 10th grades. These exams cover topics including arts, English, foreign languages, math, history and social sciences, health education, and the sciences (including Earth and Space Science and Physics.)

MCAS exams are designed to test material that is based on a well defined and documented set of standards (www.doe.mass.edu/frameworks). These standards identify material that students - and teachers - are expected to be proficient in. Several of the Earth and Space Science and physics content standards are directly related to the Sun. Examples include solar energy, the electromagnetic spectrum, the Sun/Earth/moon connections (including solar eclipses), the Doppler effect, methods of energy transfer, and magnetic fields. In addition, for all the sciences, students should have proficiency in mathematical skills as well as *Scientific Inquiry Skills*. Specifically, these skills include:

- Make observations, raise questions, and formulate hypotheses.
- Design and conduct scientific investigations.
- Analyze and interpret results of scientific investigations.
- Communicate and apply the results of scientific investigations.

In 2005 Shadia Rifai Habbal, Munir Nayfeh and I wrote a National Science Foundation grant to observe total solar eclipses in 2006, 2008, and 2009. We incorporated an education and public outreach component for which I am primarily responsible. One of my goals with our grant is to develop a graduate level course for in-service teachers that makes connections between the Sun and MCAS content areas as well as provide learning opportunities that specifically address the *Scientific Inquiry Skills*.

1.2. Highly Qualified Teachers

In order for a teacher to be highly qualified in her/his discipline, she/he must maintain her/his teaching certification and meet one of the following qualifications:

- Pass the state licensure test in subject area
- Undergraduate major or equivalent coursework in subject area
- Graduate degree in subject area

Bridgewater State College (BSC) in Bridgewater, Massachusetts, USA, is well known for its teacher preparation programs. In addition, BSC has become one of the premier centers in southeastern Massachusetts for teachers working toward becoming highly qualified. As a result, BSC is an ideal place to offer our graduate course on incorporating the Sun into the classroom. Not only do teachers learn about new materials for their classrooms, they also work on their own professional development.

2. Educational Material based on the Sun

Several educational materials related to the Sun have already been developed and are readily available - two comprehensive lists can be found at eo.nso.edu/resources.html and umbra.nascom.nasa.gov/outreach.html. My goal is to incorporate some of these materials into my course as well as develop new ones, some of which are directly related to our eclipse expeditions. All material and activities will be geared toward the content standards and skills that the teachers are responsible for helping students learn. Therefore, such a course will not only help teachers earn credits toward becoming highly qualified, but also help them prepare their students for the MCAS exam.

Many of the individual standards mentioned above have obvious connections with the Sun - e.g. the electromagnetic spectrum and Sun/moon/earth relationships. My plan is to develop material based on the Sun that directly addresses *scientific inquiry skills* and gives students some sense of what it is like to do scientific research. Following are two examples of such activities that I have used with students - one utilizing free, online data and another utilizing data taken during a total solar eclipse.

2.1. Activity 1: Sunspot patterns

This activity was done with a group of students in an introductory astronomy course during the Summer of 2006. The class was made up of 8 college and 2 high school students. My goals with this activity were to have students access and look at data, analyze and interpret their observations, and share their results in an informal setting as well as in writing. Prior to this activity, students had learned about the structure of sunspots, but not about the solar sunspot and magnetic cycles.

Students visited the website www.solarmonitor.org and studied two magnetogram images 6 years apart (one during solar minimum, the other during solar maximum.) They were asked to make and record observations of patterns in the sunspots. This particular group of students noticed a difference in the number of the spots on the two dates, the range of latitudes where sunspots were located, the different polarity of the leading spots in the northern and southern hemispheres, and that the sunspots came in pairs. Students shared their observations and made a comprehensive list of things that need to be explained by any sunspot model. This then led into a discussion of the Babcock model.

In addition, students were asked to estimate the rotational period of the Sun using two different spots - one at a high latitude during solar maximum, the other near the equator during solar minimum. Students shared their data, discussed their methods and range in their answers, and ways to reduce errors. My hope was that students would be able to observe differential rotation, but interestingly enough, there was no statistical difference in the rotational rate as a function of latitude based on these two spots. This disparity generated an excellent conversation about what happens during research, ways to check answers, and the benefits of multiple

observations and data points.

This activity hit upon nearly all of the *scientific inquiry skill* standards, and let the students work with real data and notice patterns on their own that connect with well established patterns such as the sunspot and magnetic cycles and the Maunder butterfly diagram. They also got a sense for how long it can take to analyze data, and the value of making many observations to observe trends.

2.2. Activity 2: Eclipse conditions

Before the 2002 solar eclipse in South Africa, I met with over 100 middle school students in Plymouth, Massachusetts. We discussed solar eclipses, how they happened, and what our research team hoped to learn from them. During my time with them, students asked how temperature and light intensity changed during a total eclipse. I had no quantitative answer for them, so I asked the students to hypothesize what would happen; they predicted it would get colder and much darker. Using equipment they had access to in their classrooms, we designed a simple experiment to quantify the changes during the eclipse. A temperature probe would be left in the shade to monitor the temperature throughout the eclipse. In addition, a light sensor would measure the amount of light reflected off a piece of white paper. Figure 1 shows the light probe on the day of the eclipse (the temperature probe is on the other side of the tree.)

Fig. 1. Middle School students helped design an experiment to measure changes in light intensity and temperature during totality. A light sensor attached to a tree and trained on a white piece of paper measures the reflected light throughout the eclipse. The temperature probe was kept in the shade on the other side of the tree.

When I returned from the eclipse, I brought the students their data, shown in figure 2, and we spent time together interpreting it. They noted interesting

properties like the numerous dips in light intensity (which they correctly determined were due to clouds) and the time delay in intensity and temperature minima (perhaps due to the high heat capacity of the Earth.)

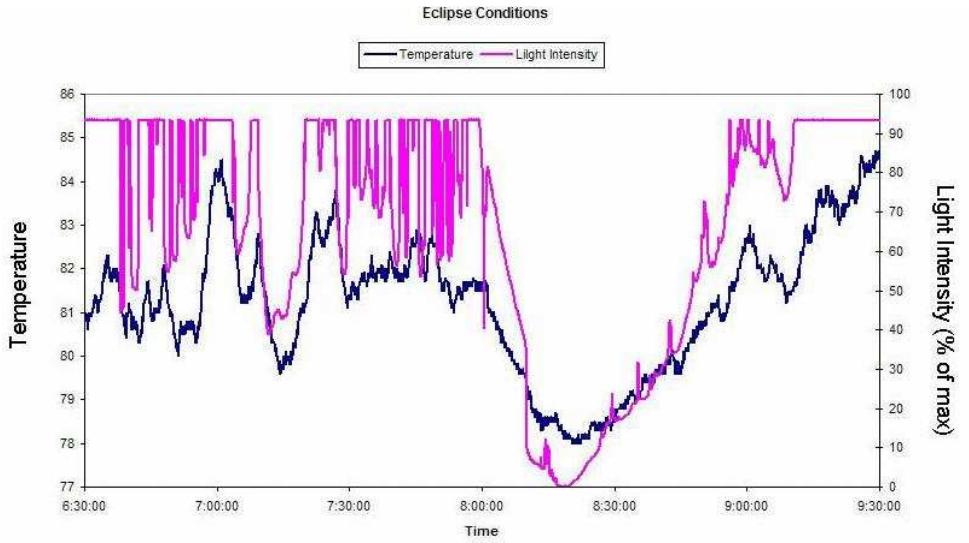


Fig. 2. Light intensity (pink) and temperature data (purple) during the 2002 eclipse. Interesting features include numerous dips in the data (due to clouds) and time delays in temperature and light intensity minima(likely due to the high heat capacity of the Earth.)

This activity engaged students in experimental design as well as data analysis and interpretation. They had a sense of ownership of the experiment and they learned how to find answers to their original questions, and check their hypotheses. Interestingly enough, when I share figure 2 with students in college level courses and groups of amateur astronomers, it always generates lively and useful discussions. This may be because the concepts behind the data and experimental setup are accessible at many levels.

3. Future work

The two examples above are only the beginning, and more work needs to be done. My goal is to offer this graduate level course during the summer of 2007. In addition to developing new materials myself, I plan to utilize the experience of the teachers and have them contribute as well. All materials will be made available to the public at www.bridgew.edu/physics.

In addition, we hope to include a middle or high school teacher in our research team during the 2008 and 2009 eclipses. The teacher would be responsible for conducting any student designed experiments as well as maintaining a journal and web page documenting the experience. Our goal is to provide the teacher with a research

experience that will ultimately impact her/his students through not just content, but also the excitement that comes with doing research on total solar eclipses.

Acknowledgments. M. Arndt would like to thank S.R. Habbal for assembling and leading such a great research team. She also thanks her colleagues in the Office of Grants and Sponsored Projects at BSC for their help administering the NSF grant supporting this work (NSF grant ATM-0450096), and T. Kling for his help in preparing figures for this publication.

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

FROM THE 1919 SOLAR ECLIPSE TO GRAVITATIONAL LENSING: ASTROPHYSICAL MANIFESTATIONS OF CURVED SPACETIME

PHILIPPE JETZER
Institute for Theoretical Physics, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Abstract. I will briefly present the history of gravitational lensing, which started with the measurement of the light deflection during the solar eclipse of May 1919 and thus confirmed the theory of general relativity. This result was then further confirmed in following solar eclipse observations. Gravitational lensing, after the first discovery of an extragalactic lens system in 1979, has rapidly developed into a very powerful astrophysical tool with several applications, among which the detection of dark matter and the discovery of extra solar planets. I will give a brief overview of the main results achieved so far.

Key words: general relativity, gravitational lensing

1. INTRODUCTION

Gravitational lensing - i.e. light deflection by gravity - has become in the last few years one of the most important fields in present day astronomy. The enormous activity in this area has mainly been driven by the considerable improvements of the observational capabilities. Due to the new wide-field cameras and telescopes which are already in place or will become operational in the near future the rate and the quality of the lensing data will increase dramatically. Gravitational lensing is independent of the nature and the physical state of the deflecting mass, therefore, it is perfectly suited to study dark matter at all scales.

Indeed, the determination of the amount and the nature of matter present in the Universe is an important problem for contemporary astrophysics and cosmology. This knowledge is directly related to the question of the fate of the Universe: will it expand forever or, after a phase of expansion, collapse again. There are several astrophysical observations which indicate that most of the matter present in the Universe is actually dark and, therefore, cannot be detected using telescopes or radio telescopes. The most recent studies lead to the conclusion that the total matter density is only about 30% of the "closure density" of the Universe: the amount of mass that would make the Universe balance between expanding forever and collapsing. Measurements based on high-redshift supernovae suggest that there is also a non-vanishing cosmological constant, such that the sum of matter density and cosmological constant implies a flat Universe (Perlmutter *et al.*, 1999).

Important evidence for the existence of large quantities of dark matter comes from the measured rotation curves of several hundreds of spiral galaxies, which imply the presence of a huge dark halo in which these galaxies are embedded. Typically, a galaxy including its halo contains ~ 10 times more dark than luminous matter, the

latter being in the form of stars and gas. There are also clear indications for the presence of important quantities of dark matter on larger scales, in particular in clusters of galaxies. This was first pointed out in 1933 by Zwicky (Zwicky, 1933). Since then, much effort has been put into the search for dark matter, the nature of which is still largely unknown.

The field of gravitational lensing is growing very rapidly and almost daily there are new results, therefore we give here only a summary of the main results. For more details see the book of Schneider et al. (1992) and the references therein.

2. HISTORICAL REMARKS

Today we know that light is bended by the presence of gravitational fields. This phenomenon is a consequence of the general relativity theory as formulated in its final version by Einstein in 1915. Within Newtonian gravity theory, since the light is massless, the trajectory of light is not affected by gravity and thus remains unchanged. Nonetheless, long before Einstein's general relativity several people thought about the possibility that light would be affected by gravity. Already, Newton in his book on Optics, published in 1704, in his first Query wrote: "Do not Bodies act upon Light at a distance, and by their action bend its Rays; and is not this action (caeteris paribus) strongest at the least distance?".

Various scientist, afterwards, made thoughts on light rays bending by massive bodies. Assuming that light has a mass one can easily compute within Newtonian mechanics how much it gets deviated by a massive body, as for instance the Sun. Moreover, it turns out that the deflection angle does not depend on the mass of the light as it drops out. Clearly, the so obtained result is only right as long as light has a mass, even if tiny. As is found using classical mechanics, a particle starting with velocity v at a large separation from the gravitating mass M is deflected by an angle α given by (assuming small deflection angles):

$$\alpha \simeq \frac{2GM}{v^2 r} \ . \tag{1}$$

From this equation, assuming that the result also holds for massless light, one obtains the "Newtonian" value for the light deflection by setting $v = c$. In particular, a light ray that grazes the surface of the Sun should be deflected by $0.''85$, as can be found by setting $r = R_\odot$ ($R_\odot$ being the radius of the Sun) and $M = M_\odot$ ($M_\odot$ being the mass of the Sun). Indeed, in 1804 the astronomer Soldner published a paper in which he computed the error induced by the light deflection on the determination of the position of stars. To that purpose he used the Newtonian theory of gravity assuming that the light is made of particles. He also estimated that a light ray which just grazes the surface of the Sun would be deflected by a value of only $0.''85$. Within general relativity this value is about twice as much, more precisely $1.''7$. The first measurement of this effect has been made during the solar eclipse of 29 May 1919 and confirmed the value predicted by general relativity (Dyson *et al.*, 1920). These observations were then repeated during following solar eclipses, in particular the ones which took place in 1922 in Australia, 1929 in Sumatra, 1936 in Russia and Japan, 1947 in Brazil and 1952 in Sudan. All results were in agreement with the prediction of general relativity.

In 1936 Einstein published a short paper in Science in which he computed the light deflection of light coming from a distant star by the gravitational field of another star (Einstein, 1936). He mentioned that if the source and the lens are perfectly aligned the image would be a ring. If instead the alignment is not perfect one would see two images with, however, a very small separation angle. Einstein also wrote: "Of course, there is no hope of observing this phenomenon". Actually, it has been found recently that Einstein made most of the calculations presented in that paper already in 1912 as can be seen on some pages of his notebook (Renn *et al.*, 1997). The recent developments of microlensing show that Einstein's conclusion, although understandable at that time, was too pessimistic. Indeed, the formulae developed by Einstein in his 1936 paper are still the basis for the description of gravitational lensing.

In the following year 1937 the swiss astronomer Zwicky wrote two short articles in Physical Review suggesting to consider galaxies as sources and lenses rather than stars as mentioned by Einstein (Zwicky, 1937). He came to the conclusion that such a configuration would have much higher chances to be seen, since the typical mass of a galaxy is several billion times higher than the mass of a single star. He argued that such configurations must almost certainly be seen. Moreover, he gave also a list of possible applications among which the possibility to better determine the total mass of galaxies, including their dark matter content.

The first gravitational lens has been discovered in 1979, when spectra were obtained of two point-like quasars which lie only about 6 arc seconds away. The spectra showed that both objects have the same redshift and are thus at the same distance. Later on also the galaxy acting as lens has been found, making it clear that the two objects are the images of the same quasar, which is lensed. Since then many other examples have been found, and in 1986 the first lensing case with a galaxy acting as source was discovered. The galaxy appears then distorted as one or more arcs. Many such systems have since then be discovered, some thanks to the Hubble space telescope.

In 1993 the first galactic microlensing events were observed, in which the source is a star in the Large Magellanic Cloud or in the galactic bulge. In the former case the lens is a compact object probably located in the galactic halo, whereas in the later case the lens is a low mass star in the galactic disk or in the bulge itself.

3. BASICS OF LENSING

The propagation of light in a curved space-time is in general a complicated problem, however, for almost all relevant applications of gravitational lensing one can assume that the geometry of the universe is described in good approximation by the Friedmann-Lemaître-Robertson-Walker metric. The inhomogeneities in the metric can be considered as local perturbations. Thus the trajectory of the light coming from a distant source can be divided into three distinct pieces. In the first one the light coming from a distant source propagates in a flat unperturbed space-time, nearby the lens the trajectory gets modified due to the gravitational potential of the lens and afterwards in the third piece the light travels again in an unperturbed space-time till it gets to the observer. The region around the lens can be described

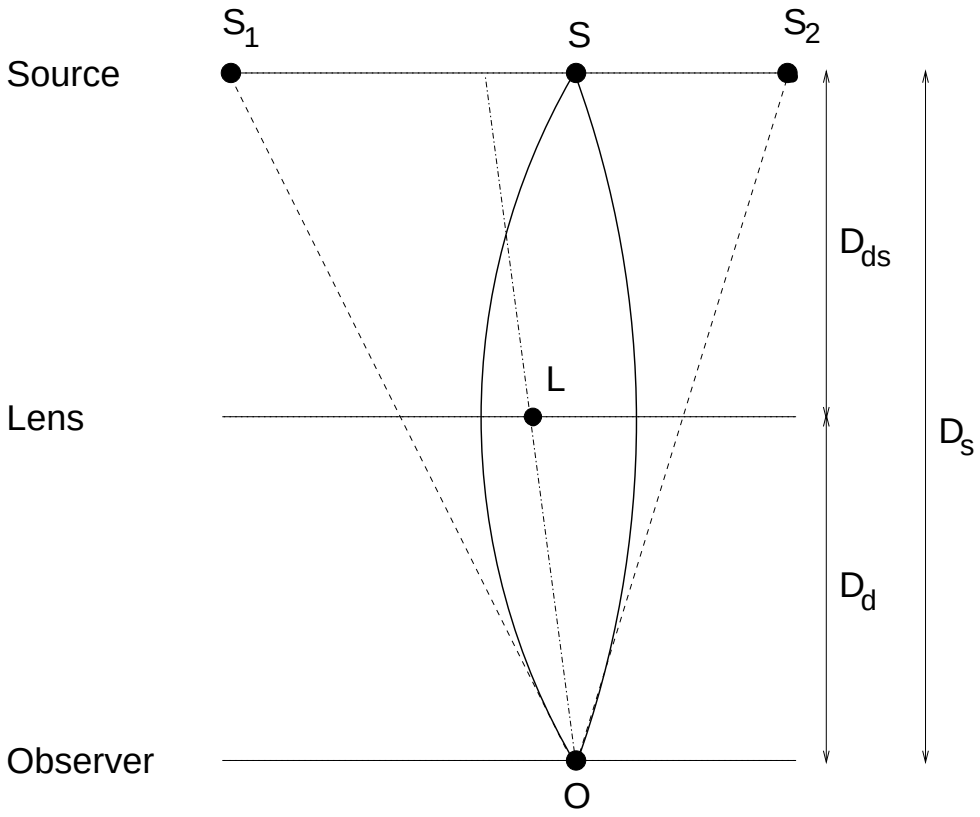


Fig. 1. Setup of a gravitational lens situation: The lens L located between source S and observer O produces two images S_1 and S_2 of the background source. D_d is the distance between the observer and the lens, D_s between the observer and the source and D_{ds} between the lens and the source.

by a flat Minkowskian space-time with small perturbations induced by the gravitational potential of the lens. This approximation is valid as long as the Newtonian potential Φ is small, which means $\mid \Phi \mid \ll c^2$ (c being the velocity of light), and if the peculiar velocity v of the lens is negligible as compared to c. These conditions are almost always fulfilled in all cases of interests for the astrophysical applications. An exception, for instance, is when the light rays get close to a black hole. We will not discuss such cases in the following.

With the above simplifying assumptions one can describe the light propagation nearby the lens in a flat space-time with a perturbation due to the gravitational potential of the lens described in first order post-Newtonian approximation. The effect of the space-time curvature on the light trajectory can be described as an effective refraction index. One can thus derive a lens equation purely geometrical optics with an effective refraction index given by:

$$n = 1 - \frac{2\Phi}{c^2} \ . \tag{2}$$

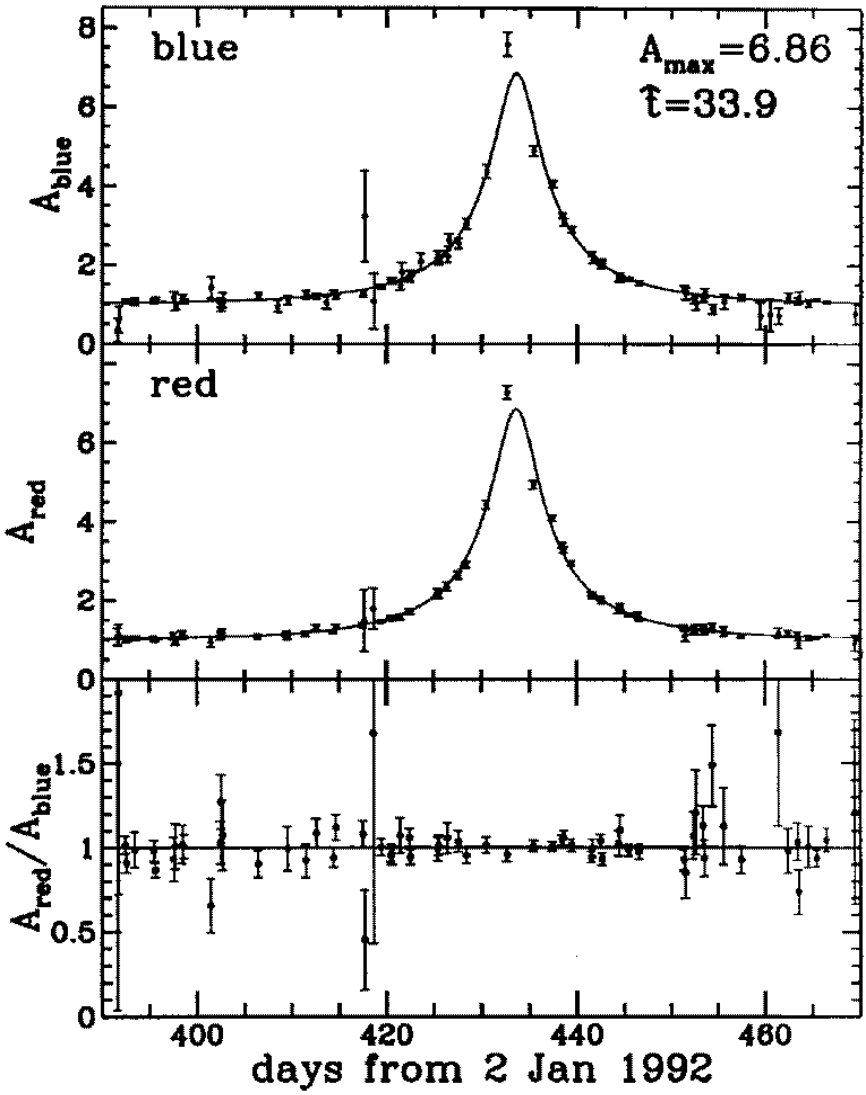


Fig. 2. Microlensing event observed by the MACHO collaboration in their first year data towards the LMC. The event lasted about 33 days. The data are shown for blue light, red light and the ratio red light to blue light, which for perfect achromaticity should be equal to 1 (from (Alcock *et al.*, 1993)).

4. MICROLENSING

There are cases in which the deflection angles are tiny, of the order of milliarcseconds or smaller, such that the multiple images are not observable. However, lensing magnifies the affected source, and since the lens and the source are moving relative

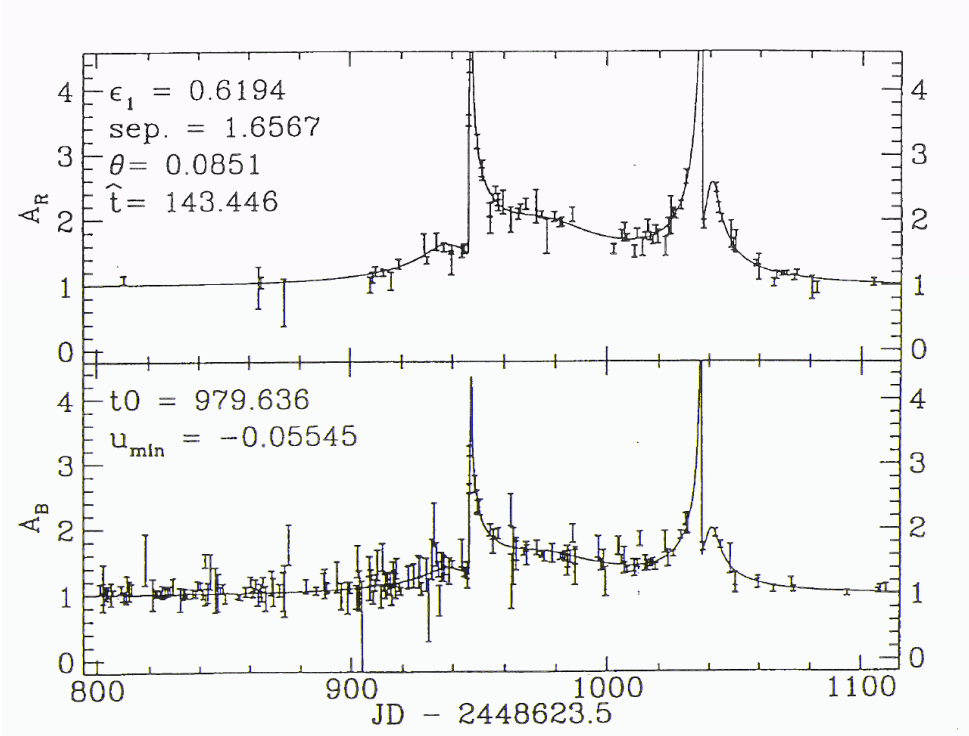


Fig. 3. Binary microlensing event towards the LMC by the MACHO collaboration (taken from the web page http://darkstar.astro.washington.edu). The two light curves correspond to observations in different colors taken in order to test achromaticity.

to each other, this can be detected as a time-variable brightness. This behaviour is referred to as gravitational microlensing, a powerful method to search for dark matter in the halo of our own Galaxy, if it consists of massive astrophysical compact halo objects (MACHOs), and to study the content of low-mass stars in the galactic disk.

The idea to use gravitational light deflection to detect MACHOs in the halo of our Galaxy by monitoring the light variability of millions of stars in the Large Magellanic Cloud (LMC) was first proposed by Paczyński (1986). Since then the field has grown very rapidly, especially since the discovery of the first microlensing events at the end of 1993 and many new applications have been suggested, including the detection of Earth like planets around stars in our Galaxy.

Since the discovery of the first microlensing events in September 1993 by monitoring millions of stars in the Large Magellanic Cloud (LMC) and in the direction of the galactic centre, several hundreds of events have been found. The still few observed events towards the LMC indicate that the halo dark matter fraction in the form of MACHOs is at most of the order of 20%, assuming a standard spherical halo model.

4.1. Microlensing towards the LMC

Microlensing allows the detection of MACHOs located in the galactic halo in the mass range $10^{-7} < M/M_{\odot} < 1$ (De Rújula *et al.*, 1991; De Rújula *et al.*, 1992), as well as MACHOs in the disk or bulge of our Galaxy (Paczyński, 1991; Griest *et al.*, 1991).

In September 1993, the French collaboration EROS (Expérience de Recherche d'Objets Sombres) (Aubourg *et al.*, 1993) announced the discovery of two microlensing candidates, and the American–Australian collaboration MACHO (for the collaboration they use the same acronym as for the compact objects) of one candidate (Alcock *et al.*, 1993) by monitoring several millions of stars in the LMC (see Fig. 2).

The MACHO team went on to report the observation of 13 to 17 events (one being a binary lensing event; see Fig. 3) by analysing their 5.7 years of LMC data (Alcock *et al.*, 2000). The inferred optical depth (which is a measure for the probability to detect a microlensing event) is $\tau = 1.2^{+0.4}_{-0.3} \times 10^{-7}$ with an additional 20% to 30% of systematic error. Correspondingly, this implies that about 20% of the halo dark matter is in the form of MACHOs with a most likely mass in the range 0.15 - 0.9 $M_{\odot}$ depending on the halo model. Moreover, it might well be that not all the MACHOs are in the halo: some could be stars in the LMC itself or located in an extended disk of our Galaxy, in which case an average mass value including all events would produce an incorrect value. These considerations show that at present, the value for the average mass as well as the fraction of halo dark matter in form of MACHOs have to be treated with care (Mancini *et al.*, 2004).

One of the events discovered was due to a lens made of two objects, namely a binary system. Such events are more rare, but their observation is not surprising; since almost 50% of the stars are double systems, it is quite plausible that MACHOs also form binary systems. The light curve is then more complicated than for a single MACHO (see Fig. 3).

EROS has also searched for very low mass MACHOs by looking for microlensing events with time scales ranging from 30 min to 7 days (Renault *et al.*, 1997). The lack of candidates in this range places significant constraints on any model for the halo that relies on objects in the range $5 \times 10^{-8} < M/M_{\odot} < 2 \times 10^{-2}$. Indeed, such objects may make up at most 20% of the halo dark matter (in the range between $5 \times 10^{-7} < M/M_{\odot} < 2 \times 10^{-3}$ at most 10%). Similar conclusions have also been reached by the MACHO group (Alcock *et al.*, 2000). A few events have also been discovered towards the Small Magellanic Cloud.

4.2. Microlensing towards other targets

To date, the MACHO, EROS and OGLE (Optical Gravitational Lensing Experiment) collaborations have found more than thousand microlensing events towards the galactic bulge, which also imply the presence of a bar in the galactic centre. They also found several events by monitoring the spiral arms of our Galaxy. These results are important for studying the structure of our Galaxy (Grenacher *et al.*, 1999). Microlensing observations towards the galactic centre turns out to be also a very powerful way to detect planetary systems in our Galaxy. In particular it allows to find planets with a mass comparable to the Earth's mass. Recently, a planet with a mass of only about five times Earth's mass has been found this way (Beaulieu

et al., 2006).

Microlensing searches towards the Andromeda galaxy (M31) have also been proposed (Crotts, 1992; Baillon *et al.*, 1993; Jetzer *et al.*, 1994). In this case, however, one has to use the so-called "pixel-lensing" method. Since the source stars are in general no longer resolvable, one has to consider the luminosity variation of a whole group of stars, which are, for instance, registered on a single pixel element of a CCD camera. This makes the subsequent analysis more difficult, however it allows to use M31 and other objects as targets. For information on the shape of the dark halo, which is presently unknown, it is important to observe microlensing in different directions. Several groups performed such observations and preliminary results suggest that about 20% of the halo dark matter could be in form of MACHOS (Calchi Novati *et al.*, 2005), however more data is needed in order to confirm this result. Such efforts are under way and possibly in few years this issue will be solved.

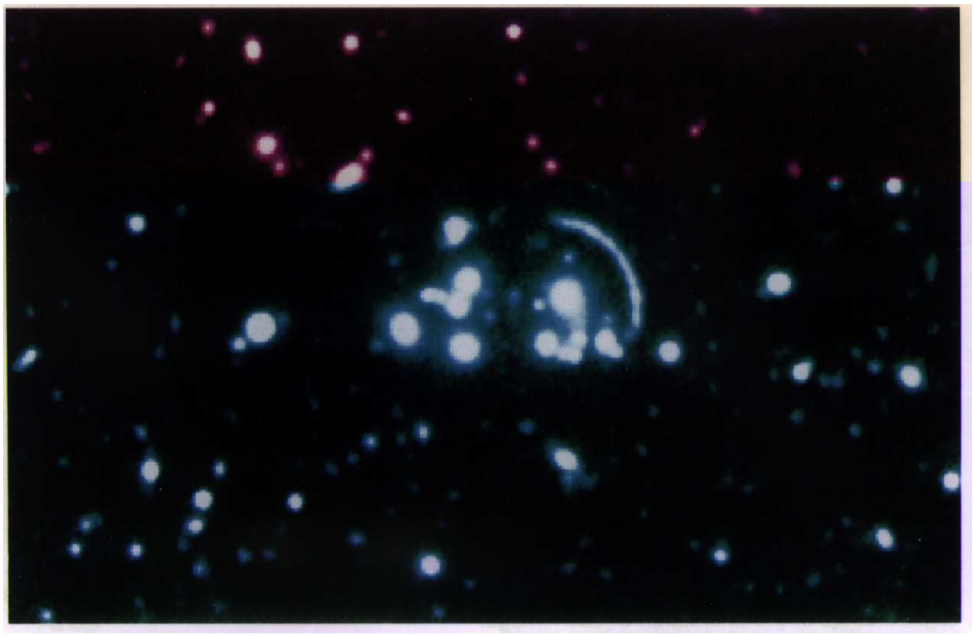


Fig. 4. Giant arc in Cl2244-02 (image from CFHT). The lensing cluster is at $z = 0.329$ and the source of the arc is a very distant field galaxy at $z = 2.238$. (Courtesy of G. Soucail, Obs. Midi–Pyrénées, ESO Messenger 69, September 1992.)

5. GALAXY CLUSTERS AS LENSES

When dealing with cosmological lenses one has to interpret all distances as *angular diameter distances*. Galaxy clusters similarly to galaxies can act as gravitational lenses for more distant galaxies (see Figs. 4 and 5). One classifies the observed lensing effects due to clusters into two types.

1) Rich centrally condensed clusters produce sometimes a giant arc (Fig. 4) when

a background galaxy turns out to be almost aligned with the cluster (*strong lensing*).

2) Every cluster produces weakly distorted images of a large number of background galaxies (*weak lensing*).

Both these cases have been observed and allow to get important information on the distribution of the matter in galaxy clusters. For the analysis of giant arcs, we have to use parameterized lens models which are fitted to the observational data. The situation is much better for weak lensing, because there now exist several parameter–free reconstruction methods of projected mass distributions from weak lensing data.

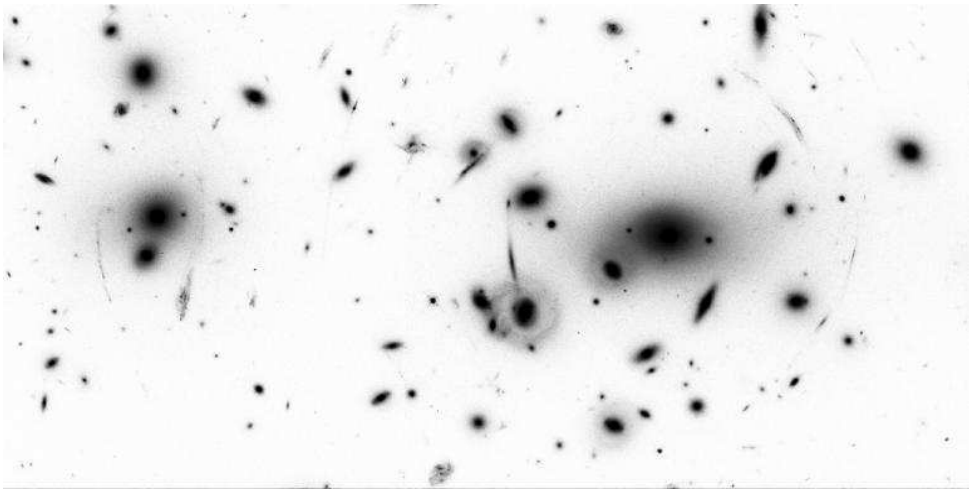


Fig. 5. Setup of a gravitational lens situation: Hubble Space Telescope image of the cluster Abell 2218. Beside arcs around the two centers of the cluster, many arclets can be seen (NASA HST Archive).

5.1. HUBBLE CONSTANT FROM TIME DELAY

As first noted by Refsdal in 1966 (Refsdal, 1966), time delay measurements can yield, in principle, the Hubble parameter H_0. Unfortunately, the use of this method requires a reliable lens model. This introduces systematic uncertainties.

Measuring the time delay is not an easy task, nonetheless for several lens systems one could get reliable observations. For instance for the double lens quasar QSO0957+561 it could be determined and turned out to be: $\Delta t = 417 \pm 3$ days. Modelings lead then for this case to a best estimate of $H_0 \simeq 61$ km s^{-1} Mpc^{-1}. For this example there are constraints for modeling the lens, nevertheless it is difficult to assess an error for the value of H_0.

Besides having the above mentioned problems, the determination of H_0 through gravitational lensing offers also some advantages as compared to the other methods. It can be directly used for large redshifts (~ 0.5) and it is independent of any other method. Moreover, it is based on fundamental physics, while other methods rely on models for variable stars (Cepheids), or supernova explosions (type II), or empirical calibrations of standard candles (Tully–Fisher distances, type I supernovae).

Finally, we notice that lensing can also lead to bounds on the cosmological constant Λ. The volume per unit redshift of the universe at high redshifts increases for a large Λ. This implies that the relative number of lensed sources for a given comoving number density of galaxies increases rapidly with Λ. This can be used to constrain Λ by making use of the observed probability of lensing. Various authors have used this method and came up with a limit $\Omega_\Lambda \leq 0.6$ for a universe with $\Omega_0 + \Omega_\Lambda = 1$. It remains to be seen whether such bounds, based on lensing statistics, can be improved.

6. Conclusions

Although being rather young observational techniques, gravitational lensing and microlensing, which had their beginning in observations made during the 1919 solar eclipse, have already enabled us to make substantial progress on various topics such as dark matter and the search of planets in our Galaxy. The prospects for further contributions to solve important astrophysical problems look very bright.

Acknowledgments. I would like to thank the organizers of the meeting for the invitation to give a talk and for the unforgettable experience.

References

Alcock, C., et al.: 1993, Nature **365**, 621; 1995, Astrophys. J. **445**, 133
Alcock, C., et al.: 2000, Astrophys. J. **542**, 281
Aubourg, E. et al.: 1993, Nature **365**, 623
Baillon, P., Bouquet, A., Giraud-Héraud, Y., and Kaplan, J.: 1993, Astron. and Astrophys. **277**, 1
Beaulieu, J.P., et al.: 2006, Nature **439**, 437..
Calchi Novati, S., et al.: 2005, Astron. and Astrophys. **443**, 911
Crotts, A.P.: 1992, Astrophys. J. **399**, L43
De Rújula, A., Jetzer, Ph., and Massó, E.: 1991, Mont. Not. R. Astr. Soc. **250**, 348
De Rújula, A., Jetzer, Ph., and Massó, E.: 1992, Astron. and Astrophys. **254**, 99
Dyson, F.W., Eddington, A.S., and Davidson, C.R.: 1920, Mem. Royal Astronomical Society **62**, 291
Einstein, A.: 1936, Science **84**, 506
Grenacher, L., Jetzer, Ph., Strässle, M., and De Paolis, F.: 1999, Astron. and Astrophys. **351**, 775
Griest, K. et al.: 1991, Astrophys. J. **372**, L79
Jetzer, Ph.: 1994, Astron. and Astrophys. **286**, 426
Mancini, L., Calchi Novati, S., Jetzer, Ph., and Scarpetta, G.: 2004, Astron. and Astrophys. **427**, 61
Paczyński, B.: 1986, Astrophys. J. **304**, 1
Paczyński, B.: 1991, Astrophys. J. **371**, L63
Perlmutter, S. et al.: 1999 , Astrophys. J. **517**, 565
Refsdal, S.: 1966, Mont. Not. R. Astr. Soc. **134**, 315
Renault, C. et al.: 1997, Astron. and Astrophys. **324**, L69
Renn, J., Sauer, T., and Stachel, J.: 1997, Science **275**, 184
Schneider, P., Ehlers, J., and Falco, E.E.: 1992, *Gravitational Lensing*, (Springer Verlag, Berlin 1992)
Zwicky, F.: 1933, Helv. Phys. Acta **6**, 110
Zwicky, F.: 1937, Phys. Rev. **51**, 290; Phys. Rev. **51**, 679

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

A GENTLE INTRODUCTION TO THE PHYSICS OF SPECTRAL LINE POLARIZATION

JAVIER TRUJILLO BUENO
Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain

Abstract. This paper presents a pedagogical introduction to the physical mechanisms that produce spectral line polarization in stellar atmospheres, emphasizing the importance of developing reliable diagnostic tools that take proper account of the Zeeman and Paschen-Back effects, scattering polarization and the Hanle effect. Only in this way may we hope to investigate the complexity of solar and stellar magnetic fields in a parameter domain that goes from field intensities as low as one milligauss to many thousands of gauss.

Key words: polarization, scattering, solar and stellar magnetic fields

1. Introduction

The state of polarization of a quasi-monochromatic beam of electromagnetic radiation can be conveniently characterized in terms of four quantities that can be measured by furnishing our telescopes with a polarimeter. Such observables are the four Stokes parameters (I, Q, U, V), which were formulated by Sir George Stokes in 1852 and introduced into astrophysics by the Nobel laureate Subrahmanyan Chandrasekhar in 1946. The Stokes $I(\lambda)$ profile represents the intensity as a function of wavelength, Stokes $Q(\lambda)$ the *intensity difference* between vertical and horizontal linear polarization, Stokes $U(\lambda)$ the intensity difference between linear polarization at $+45^\circ$ and -45°, while Stokes $V(\lambda)$ the intensity difference between right- and left-handed circular polarization (cf. Born & Wolf, 1994). Note that the definition of the Stokes Q and U parameters requires first choosing a reference direction for $Q > 0$ in the plane perpendicular to the direction of propagation.

Let us now introduce the most important mechanisms that induce (and modify) polarization signatures in the spectral lines that originate in stellar atmospheres: the Zeeman and Paschen-Back effects, scattering processes and the Hanle effect.

2. The Zeeman effect

As illustrated in Figure 1, the Zeeman effect requires the presence of a magnetic field which causes the atomic and molecular energy levels to split into different magnetic sublevels characterized by their magnetic quantum number M (Condon & Shortley, 1935). Each level of total angular momentum J splits into $(2J + 1)$ sublevels, the splitting being proportional to the level's Landé factor, g_J, and to the magnetic field strength. As a result, a spectral line between a lower level with (J_l, g_l) and an upper level with (J_u, g_u) is composed of several individual components whose frequencies are given by $\nu_{J_l M_l}^{J_u M_u} = \nu_0 + \nu_L(g_u M_u - g_l M_l)$, where ν_0 is the frequency

of the line in the absence of magnetic fields and $\nu_L = 1.3996{\times}10^6 B$ is the Larmor frequency (with B the magnetic field strength expressed in gauss). In particular, a line transition with $J_l = 0$ and $J_u = 1$ has three components (see Fig. 1): one π component centered at ν_0 (or at λ_0), one $\sigma_{\rm red}$ component centered at $\nu_0 - g_u \nu_L$ (or at $\lambda_0 + g_u \Delta\lambda_B$), and one $\sigma_{\rm blue}$ component centered at $\nu_0 + g_u \nu_L$ (or at $\lambda_0 - g_u \Delta\lambda_B$), where $\Delta\lambda_B = 4.6686{\times}10^{-13}\lambda_0^2 B$ (with λ_0 in Å and B in gauss).

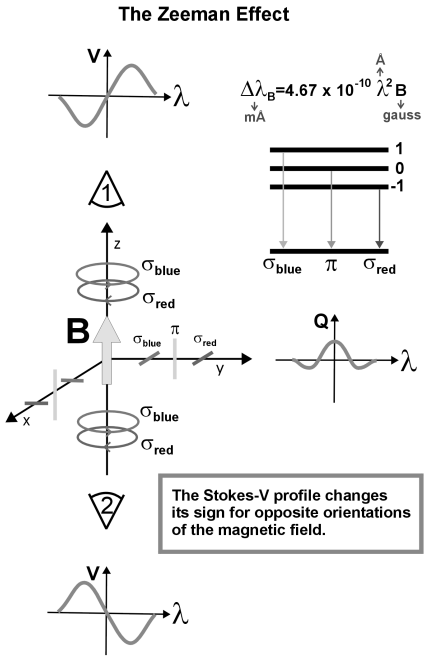


Fig. 1. The oscillator model for the Zeeman effect indicating the characteristic shapes of the circular and linear polarization profiles as generated locally via the emission process. It is important to note that the Stokes $V(\lambda)$ profile changes its sign for opposite orientations of the magnetic field vector, while the Stokes $Q(\lambda)$ profile reverses sign when the transverse field component is rotated by $\pm 90°$.

The important point to remember is that the polarization signals produced by the Zeeman effect are caused by the *wavelength shifts* between the π ($\Delta M = M_u - M_l = 0$) and $\sigma_{b,r}$ ($\Delta M = \pm 1$) transitions. Such wavelength shifts are also the physical origin of the spectral line polarization induced by the Paschen-Back effect discussed below in Section 7, since the only difference with respect to the linear Zeeman effect theory considered here lies in the calculation of the positions and strengths of the various π and σ components.

The Zeeman effect is most sensitive in circular polarization (quantified by the Stokes V parameter), with a magnitude that for not too strong fields scales with the ratio between the Zeeman splitting and the width of the spectral line (which is very much larger than the natural width of the atomic levels!), and in such a way that the emergent Stokes $V(\lambda)$ profile changes its sign for opposite orientations

of the magnetic field vector. This so-called *longitudinal* Zeeman effect responds to the line-of-sight component of the magnetic field. Accordingly, if we have a perfect cancellation of mixed magnetic polarities within the spatio-temporal resolution element of the observation, the measured circular polarization would be exactly zero *if* the thermodynamic and dynamic properties of the mixed magnetic components are similar. The antisymmetric shape of the Stokes $V(\lambda)$ profiles illustrated in Fig. 1 can be easily understood by noting the expression of the Stokes-V component of the emission vector:

$$\epsilon_V = \frac{h\nu}{4\pi} N_u A_{ul} \frac{1}{2} [\phi_{\rm red} - \phi_{\rm blue}] \cos\theta, \tag{1}$$

where θ is the angle between the magnetic field vector and the line of sight, A_{ul} the Einstein coefficient for the spontaneous emission process, N_u the number of atoms per unit volume in the upper level of the line transition under consideration, while $\phi_{\rm red}$ and $\phi_{\rm blue}$ are profiles that result from the superposition of the Voigt functions corresponding to each individual component. $\phi_{\rm red}$ is displaced to the red side of the central wavelength λ_0, and $\phi_{\rm blue}$ to the blue side. For instance, for the particular case of a line transition with $J_l = 0$ and $J_u = 1$, $\phi_{\rm red}$ is a Voigt profile centered at $\lambda_0 + g_u \Delta\lambda_B$ and $\phi_{\rm blue}$ a Voigt profile centered at $\lambda_0 - g_u \Delta\lambda_B$.

In contrast, the *transverse* Zeeman effect responds to the component of the magnetic field perpendicular to the line of sight, but produces linear polarization signals (quantified by the Stokes Q and U parameters) that are normally below the noise level of present observational possibilities for intrinsically weak fields (typically B < 100 gauss for solar spectropolarimetry). The Stokes Q and U profiles induced by the Zeeman effect at a given point in a magnetized plasma have a three-lobe shape which is also illustrated in Fig. 1. This characteristic shape can be easily understood by noting that the expression of the Stokes-Q component of the emission vector is:

$$\epsilon_Q = \frac{h\nu}{4\pi} N_u A_{ul} \frac{1}{2} [\phi_\pi - (\frac{\phi_{\rm red} + \phi_{\rm blue}}{2})] \sin^2\theta \cos 2\chi, \tag{2}$$

where χ is the angle that the projection line of the magnetic field vector on the plane perpendicular to the direction of propagation forms with the reference direction chosen for $Q > 0$.

Figure 2 shows an interesting example of Stokes profiles produced by the solar atmospheric plasma. The top panel is a section of the Fraunhofer spectrum between 4602 Å and 4610 Å showing the familiar absorption lines corresponding to several chemical elements. The remaining panels give the fractional polarizations $X(\lambda)/I(\lambda)$ (with $X = Q, U, V$). The spectrograph slit was placed parallel to and 2.5 arcsec inside the limb (at $\mu = 0.07$, where μ is the cosine of the heliocentric angle), such that half of the slit covers a significantly magnetized region, while the other half lies outside it. In the magnetically active region (which corresponds to the lower half of each of the four panels of Fig. 2) we see the characteristic signatures of the Zeeman effect. The V/I panel shows the typical antisymmetric signature of the longitudinal Zeeman effect with a positive and a negative lobe for each spectral line, while in the lower half of the panels for Q/I and U/I we see the typical symmetric signature of the transverse Zeeman effect with two lobes in the wings of opposite sign to the central lobe. Interestingly, as soon as we go outside the facular region (see the upper

half of each of the four panels in Fig. 2) we see that the amplitude of the circular polarization is significantly reduced in all atomic lines, while practically the only existing linear polarization signal is the Q/I peak corresponding to the Sr I line at 4607 Å . However, the shape of this Q/I profile is Gaussian-like, suggesting that it is not produced by the transverse Zeeman effect. If it does not result from the Zeeman effect, what, then, could its physical origin be?

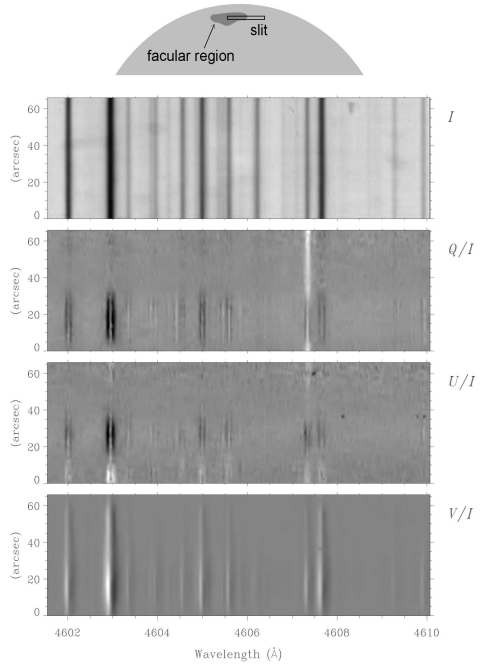


Fig. 2. Spectropolarimetric observation close to the edge of the solar disk with half of the spectrograph slit crossing a moderately magnetized facular region. Note that while the characteristic signature of the longitudinal Zeeman effect is present at all spatial points along the slit, the signature of the transverse Zeeman effect disappears as soon as one goes outside the facular region. Interestingly, the only spectral line which shows linear polarization outside the facular region is the Sr I line at 4607 Å with a Q/I shape that has nothing to do with the transverse Zeeman effect. This spectropolarimetric observation was obtained by Stenflo (2002).

3. Atomic level polarization

The amplitudes of polarization signals induced by the Zeeman effect are very small when the Zeeman splitting is a very small fraction of the spectral line width. If there is no Zeeman splitting, there is no wavelength shift between the π and σ transitions, and there is no measurable polarization because the polarizations of such components cancel out. However, it is easy to see that this is only true if the populations of the magnetic sublevels pertaining to the lower and/or upper levels of the spectral line under consideration are assumed to be identical.

To this end, consider the case of a line transition with $J_l = 0$ and $J_u = 1$ and choose the quantization axis of total angular momentum along the solar radius vector through the observed point. Assume that the population of the upper-level magnetic sublevel with $M_u = 0$ is smaller than the populations of the magnetic sublevels with $M_u = \pm 1$. As a result, even in the absence of a magnetic field (zero Zeeman splitting), we can have a non-zero linear polarization signal, simply because the number of σ transitions per unit volume and time will be larger than the ensuing number of π transitions. For example, for the case of magnetic field oriented along the solar radius vector through the observed point, a more general expression for the Stokes-Q component of the emission vector would be[1]

$$\epsilon_Q = \frac{h\nu}{4\pi} A_{ul} \frac{3}{2} [\, \rho_1(0,0)\phi_\pi \, - \, (\, \frac{\rho_1(-1,-1)\phi_{\rm red} \, + \, \rho_1(1,1)\phi_{\rm blue}}{2} \,) \,] \sin^2\theta, \qquad (3)$$

where θ is the angle between the quantization axis of total angular momentum (chosen here along the solar radius vector through the observed point) and the line of sight, while $\rho_{J_u}(M,M)$ is the population of the upper-level sublevel with magnetic quantum number M. This expression shows clearly that in the absence of a magnetic field (i.e., when $\phi_\pi = \phi_{\rm red} = \phi_{\rm blue}$ because their central wavelengths coincide at λ_0 for $B = 0$ gauss) the Stokes-Q profile is equal to the difference between two Voigt profiles: one (resulting from the π transition) centered at λ_0 and an extra one (resulting from the σ transitions) centered also at λ_0 but of greater amplitude. This is precisely the explanation of the curious linear polarization of the Sr I line seen in Fig. 2, as observed in "quiet" regions close to the edge of the solar disk. Given that this spectral line has $J_l = 0$, its linear polarization is totally due to the *selective emission* processes resulting from the population imbalances of the upper level.

On the other hand, it is very important to understand that whenever the Zeeman splitting is a negligible fraction of the spectral line width, spectral lines with $J_l = 1$ and $J_u = 0$ can produce linear polarization only if there exist population imbalances among the magnetic sublevels of their lower-level. If this is the case, then linear polarization can be generated via the *selective absorption* of polarization components resulting from the population imbalances of the lower level (Trujillo Bueno & Landi Degl'Innocenti, 1997; Trujillo Bueno, 1999, 2001, 2003*a*; Trujillo Bueno et al., 2002*a*). The same applies to $J_l = 3/2 \rightarrow J_u = 1/2$ transitions, like the $\lambda 8662$ Å line of the Ca II IR triplet (Manso Sainz & Trujillo Bueno, 2003*a*). Interestingly, lower-level atomic polarization and the ensuing selective absorption mechanism (i.e., 'zero-field' dichroism) is the physical origin of the 'enigmatic' signals of the linearly-polarized solar limb spectrum (or *second solar spectrum*) which has been discovered recently using novel polarimeters that allow the detection of very low amplitude polarization signals (with $10^{-6} < Q/I < 10^{-3}$; see Stenflo & Keller, 1997; Stenflo et al., 2000; Gandorfer, 2000, 2002).

In summary, spectral line polarization can be produced by the mere presence of *atomic level polarization*, i.e., by the existence of population imbalances among the sublevels pertaining to the upper and/or lower atomic levels involved in the line

[1] I have chosen here the positive reference direction for $Q > 0$ such that $\cos 2\chi = 1$.

transition under consideration.[2] Upper-level polarization produces selective emission of polarization components (i.e., the emitted light is polarized, even in the absence of a magnetic field), while lower-level polarization causes selective absorption of polarization components (i.e., the transmitted beam is polarized, even in the absence of a magnetic field).

4. Anisotropic radiation pumping

What is the key physical mechanism that induces atomic level polarization in a stellar atmosphere? The answer lies in the *the anisotropic illumination* of the atoms, which produces *atomic alignment*. This is easy to understand by considering the academic case of a unidirectional unpolarized light beam that illuminates a gas of two-level atoms with $J_l = 0$ and $J_u = 1$ and that is propagating along the direction chosen as the quantization axis of total angular momentum. Since these atoms can only absorb ± 1 units of angular momentum from the light beam, only transitions corresponding to $\Delta M = \pm 1$ are effective, so that no transitions occur to the $M = 0$ sublevel of the upper level. Thus, in the absence of any relaxation mechanisms, the upper-level sublevels with $M = 1$ and $M = -1$ would be more populated than the $M = 0$ sublevel and the alignment coefficient $\rho^2_0(J_u = 1) = (N_1 - 2N_0 + N_{-1})/\sqrt{6}$ would have a positive value.

Upper-level selective population pumping occurs when some upper state sublevels have more chance of being populated than others. On the contrary, as illustrated in Fig. 3, lower-level selective depopulation pumping occurs when some lower state sublevels absorb light more strongly than others. As a result, an excess population tends to build up in the weakly absorbing sublevels (Kastler, 1950; Happer, 1972; Trujillo Bueno & Landi Degl'Innocenti, 1997; Trujillo Bueno, 1999, 2001, 2003*a*; Manso Sainz & Trujillo Bueno, 2003*a*). It is also important to note that line transitions between levels having other total angular momentum values (e.g., $J_l = J_u = 1$) permit the transfer of atomic polarization between both levels via a process called *repopulation pumping* (e.g., lower-level atomic polarization can result simply from the spontaneous decay of a polarized upper level; see Trujillo Bueno et al., 2002*b*). The presence of a magnetic field is not necessary for the operation of such optical pumping processes, which can be particularly efficient in creating atomic polarization if the depolarizing rates from elastic collisions are sufficiently low. Figure 4 illustrates the type of anisotropic illumination in the outer layers of a stellar atmosphere.

5. The Hanle effect: classical description

The Hanle effect is the modification of the atomic-level polarization (and of the ensuing observable effects on the emergent Stokes profiles) caused by the action of a

[2] There are two possible types of atomic level polarization: alignment and orientation. Atomic alignment is a condition of population imbalances between the Zeeman substates of a level, such that the total population of substates with different values of $|M|$ are different. One speaks instead of atomic orientation when, for a given value of $|M|$, the substates labeled by M and -M have different populations. Moreover, the concept of atomic polarization includes also the possibility of quantum interferences (or coherences) among the magnetic sublevels of each J-level, and even among those belonging to different J-levels.

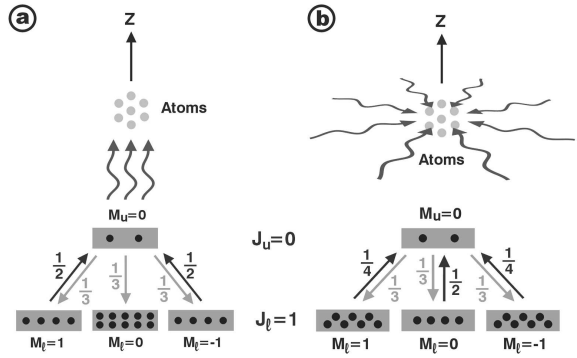


Fig. 3. Illustration of the atomic polarization that is induced in the lower level of a two-level atom (with $J_l = 1$ and $J_u = 0$) by two types of anisotropic illuminations (**a** and **b**). The incident radiation field is assumed to be unpolarized and with axial symmetry around the vertical direction, which is our choice here for the quantization axis of total angular momentum. In both cases, an excess population tends to build up in the weakly absorbing sublevels. Note that the alignment coefficient of the lower level (i.e. $\rho^2_0 = (N_1 - 2N_0 + N_{-1})/\sqrt{6}$, N_i being the populations of the magnetic sublevels) is negative in case (**a**) (where the incident beam is parallel to the quantization axis), but positive in case (**b**) (where the incident beams are perpendicular to the quantization axis).

magnetic field *inclined* with respect to the symmetry axis of the pumping radiation field. The basic formula to estimate the magnetic field intensity, $B_{\rm H}$ (measured in gauss), sufficient to produce a sizable change in the atomic level polarization results from equating the Zeeman splitting with the natural width (or inverse lifetime) of the energy level under consideration:

$$B_H = 1.137{\times}10^{-7}/(t_{\rm life}\, g_J)\,, \tag{4}$$

where g_J and $t_{\rm life}$ are, respectively, the Landé factor and the level's lifetime (in seconds), which can be either the upper or the lower level of the chosen spectral line. This formula provides a reliable estimation only when radiative transitions dominate completely the atomic excitation. If elastic and/or inelastic collisions are also efficient, then the critical field increases, since it turns out to be approximately given by (Trujillo Bueno, 2003*a*)

$$B \approx \frac{1+\delta(1-\epsilon)}{1-\epsilon}\, B_H\,, \tag{5}$$

where $\delta = D/A_{ul}$ quantifies the rate of elastic (depolarizing) collisions in units of the Einstein A_{ul}-coefficient, and $\epsilon = C_{ul}/(C_{ul} + A_{ul})$ is the probability that a de-excitation event is caused by collisions (with C_{ul} the rate of inelastic collisional transitions between the upper level "u" and the lower level "l"). The application of this basic equation to the upper and lower levels of typical spectral lines shows that the Hanle effect may allow us to diagnose solar and stellar magnetic fields having intensities between at least one milligauss and a few hundred gauss, i.e., in a parameter domain that is very hard to study via the Zeeman effect alone.

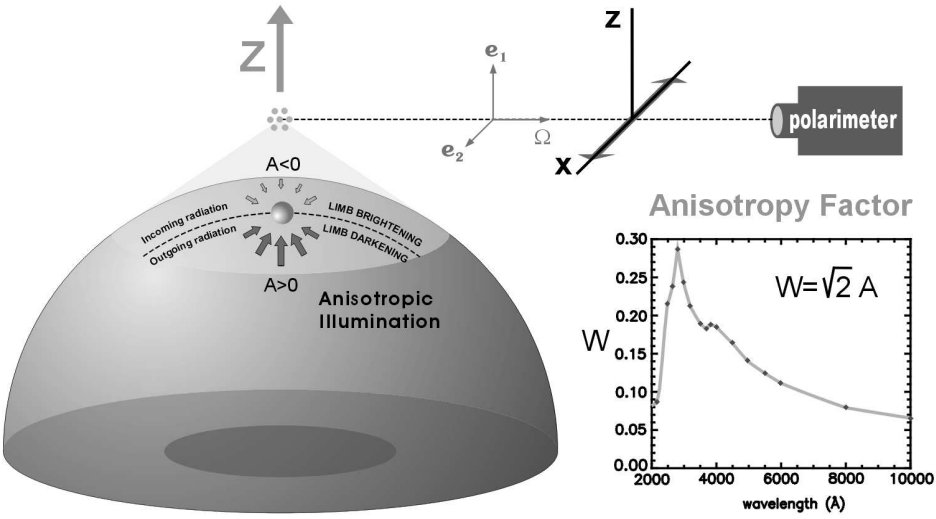


Fig. 4. Anisotropic illumination of the outer layers of a stellar atmosphere, indicating that the outgoing continuum radiation (which shows limb darkening) is predominantly *vertical* while the incoming radiation (which shows limb brightening) is predominantly *horizontal* . The figure also illustrates the type of anisotropic illumination experienced by atoms situated at a given height above the visible 'surface' of the star, including the polarization analysis of the scattered beam at 90°. The 'degree of anisotropy' of the incident radiation field is quantified by $A = J^2_0/J^0_0$, where J^0_0 is the familiar mean intensity and $J^2_0 \approx \oint \frac{d\vec{\Omega}}{4\pi} \frac{1}{2\sqrt{2}}(3\mu^2-1) I_{\nu,\vec{\Omega}}$ (with $I_{\nu,\vec{\Omega}}$ the Stokes-I parameter as a function of frequency ν and direction $\vec{\Omega}$, while $\mu = \cos\theta$, with θ the polar angle with respect to the Z-axis). The possible values of the 'anisotropy factor' $W = \sqrt{2}\,A$ vary between $W = -1/2$, for the limiting case of illumination by a purely horizontal radiation field without any azimuthal dependence (case b of Fig. 3), and $W = 1$ for purely vertical illumination (case a of Fig. 3). It is important to point out that the larger the 'anisotropy factor' the larger the fractional atomic polarization that can be induced, and the larger the amplitude of the emergent linear polarization. We choose the positive direction for the Stokes-Q parameter along the X-axis, i.e. along the perpendicular direction to the stellar radius vector through the observed point. The inset shows the wavelength dependence of the anisotropy factor corresponding to the center to limb variation of the observed solar continuum radiation. Note that in this case the maximum anisotropy factor occurs around 2800 Å, i.e., very near the central wavelength of the k line of Mg II, whose polarization may contain valuable information on the magnetic fields of the transition region from the chromosphere to the 10^6 K solar coronal plasma.

In order to clarify that, depending on the scattering geometry, the Hanle effect can either destroy or create linear polarization in spectral lines, let us consider scattering processes in a $J_l = 0 \rightarrow J_u = 1$ line transition for the following two geometries: 90° scattering and forward scattering.

5.1. 90° SCATTERING

Figure 5 illustrates the 90° scattering case, in the absence and in the presence of a magnetic field. For this geometry the largest polarization amplitude occurs for the zero field reference case, with the direction of the linear polarization as indicated in the top panel (i.e, perpendicular to the scattering plane).

The two lower panels illustrate what happens when the scattering processes take place in the presence of a magnetic field pointing (a) towards the observer (left panel) or (b) away from him/her (right panel). In both situations the polarization amplitude is reduced with respect to the previously discussed unmagnetized case. Moreover, the direction of the linear polarization is rotated with respect to the zero field case. Typically, this rotation is counterclockwise for case (a), but clockwise for case (b)[3]. Therefore, when opposite magnetic polarities coexist within the spatio-temporal resolution element of the observation the direction of the linear polarization is like in the top panel of Fig. 5, simply because the rotation effect cancels out. However, the polarization amplitude is indeed reduced with respect to the zero field reference case, which provides an "observable" that can be used for obtaining empirical information on hidden, mixed polarity fields at subresolution scales in the solar atmosphere (Stenflo, 1982; Trujillo Bueno et al., 2004).

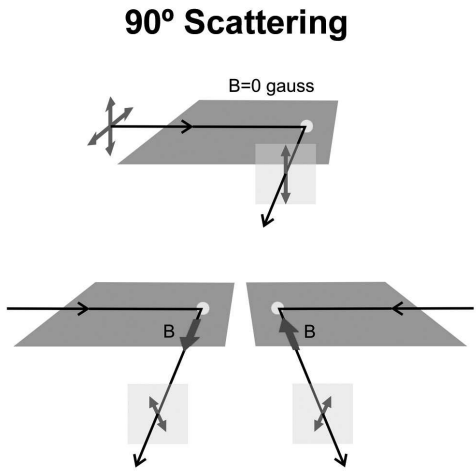


Fig. 5. The 90° scattering case in the absence (top panel) and in the presence (bottom panels) of a deterministic magnetic field.

5.2. Forward scattering

Figure 6 illustrates the case of forward scattering, in the absence and in the presence of a magnetic field. In this geometry we have *zero* polarization for the unmagnetized reference case, while the largest linear polarization (oriented along the direction of the external magnetic field) is found for "sufficiently strong" fields (i.e., for a magnetic strength such that the ensuing Zeeman splitting is much larger than the level's natural width).

In other words, in the presence of an *inclined* magnetic field that breaks the symmetry of the scattering polarization problem, forward scattering processes can produce measurable linear polarization signals in spectral lines (Trujillo Bueno, 2001).

[3] This occurs when the Landé factor, g_L, of the transition's upper level is positive, while the opposite behavior takes place if $g_L<0$.

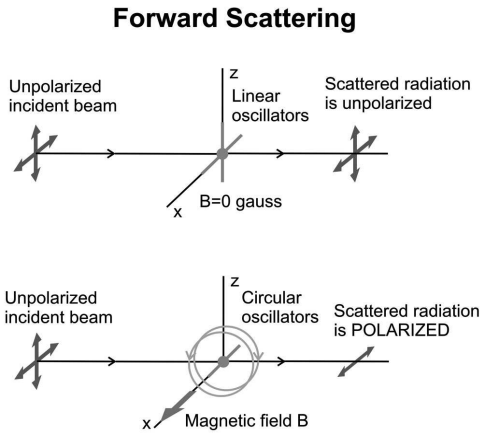


Fig. 6. The forward scattering case, in the absence (top panel) and in the presence (bottom panel) of a deterministic magnetic field.

In this case, the linear polarization is created by the Hanle effect, a physical phenomenon that has been clearly demonstrated via spectropolarimetry of solar coronal filaments in the He I 10830 Å multiplet (Trujillo Bueno et al., 2002*a*).

6. The Hanle effect: quantum description

As seen in Figs. 5 and 6, the classical description of the Hanle effect is relatively easy to understand (cf., Landi Degl'Innocenti & Landolfi, 2004). In what follows, I provide a very brief explanation within the framework of quantum mechanics, since this is required for a deeper physical understanding of this fascinating effect which has found so many interesting applications in physics (see Moruzzi & Strumia, 1991).

To that end, we need to recall first the concept of quantum coherence ($\rho_J(M, M')$) between different magnetic sublevels M and M' pertaining to each J-level. We say that the quantum coherence $\rho_J(M, M')$ is non-zero when the wave function presents a well defined phase relationship between the pure quantum states $|JM\rangle$ and $|JM'\rangle$. It is actually very common to find non-zero coherences while describing the excitation state of an atomic or molecular system under the influence of a pumping radiation field. Let us again consider a two-level atom with $J_l = 0$ and $J_u = 1$ that is being irradiated by an *unpolarized* radiation beam. In the absence of magnetic fields, all coherences of the upper level are zero *if* the quantization axis of total angular momentum is chosen along the symmetry axis of the pumping radiation beam. The same happens if a magnetic field is aligned with the quantization axis and this axis coincides with the symmetry axis of the radiation field that 'illuminates' the atomic system. This is because unpolarized radiation propagating along the quantization axis can only produce *incoherent* excitation of the upper-level sublevels with $M = \pm 1$.[4] If we now rotate the original reference system so that the new quantization

[4] Note that an unpolarized radiation beam may be considered as the incoherent superposition of right-handed and left-handed circular polarization.

axis for total angular momentum forms a non-zero angle with the symmetry axis of the radiation field, then non-zero coherences arise in this new reference system, even in the absence of a magnetic field. As shown below in Eq. (6), a magnetic field will relax such quantum coherences.

We thus see that the most general description of the excitation state of a J-level requires $(2J+1)^2$ quantities: the individual populations ($\rho_J(M,M)$) of the $(2J+1)$ sublevels and the degree of quantum coherence between each pair of them ($\rho_J(M,M')$). These quantities are nothing but the diagonal and non-diagonal elements of the atomic density matrix associated with the J-level, as given by the standard representation. Alternatively, we can use the multipole components (ρ^K_Q) of the atomic density matrix, which are given by linear combinations of $\rho_J(M,M')$. The ρ^K_Q elements with $Q=0$ are *real* numbers given by linear combinations of the populations of the various Zeeman sublevels corresponding to the level of total angular momentum J. The total population of the atomic level J is proportional to $\sqrt{2J+1}\rho^0_0(J)$, while the population imbalances among the Zeeman sublevels are quantified by ρ^K_0 (e.g., $\rho^2_0(J=1)=(N_1-2N_0+N_{-1})/\sqrt{6}$ and $\rho^1_0(J=1)=(N_1-N_{-1})/\sqrt{2}$). However, the ρ^K_Q elements with $Q\neq 0$ are *complex* numbers given by linear combinations of the *coherences* between Zeeman sublevels whose magnetic quantum numbers differ by Q (e.g., $\rho^2_2(J=1)=\rho(1,-1)$). These multipole components of the atomic density matrix provide the most useful way of quantifying, at the atomic level, the information we need for calculating the *sources* and *sinks* of polarization. Thus, the ρ^0_0 elements produce the dominant contribution to the Stokes I parameter, the ρ^1_Q elements (the orientation components) affect the circular polarization, while the ρ^2_Q elements (the alignment components) contribute to the linear polarization signals.

The Hanle effect can be suitably summarized by the following equation, which is valid for the case of a two-level model atom with unpolarized ground level (Landi Degl'Innocenti & Landolfi, 2004):

$$\rho^K_Q(J_u) \;=\; \frac{1}{1+iQ\Gamma_u}[\rho^K_Q(J_u)]_{B=0}, \tag{6}$$

where $\Gamma_u = 8.79\times10^6\ B\ g_{J_u}/A_{ul}$ and $[\rho^K_Q(J_u)]_{B=0}$ are the ρ^K_Q elements for the non-magnetic case defined in the reference frame in which the quantization axis is aligned with the magnetic field vector. This equation shows clearly that *in the magnetic field reference frame* the population imbalances (i.e., the ρ^K_Q elements with $Q=0$) are unaffected by the magnetic field, while the ρ^K_Q elements with $Q\neq 0$ are *reduced* and *dephased* with respect to the non-magnetic case. The important point to remember is that the Hanle effect modifies the emergent Stokes profiles because the polarization of the light that is emitted and/or absorbed at each point within the astrophysical plasma under consideration depends sensitively on the local values of the ρ^K_Q elements along the line of sight.

Finally, it is interesting to mention that the Hanle effect (Hanle, 1924) played a fundamental role in the development of quantum mechanics because it led to the introduction and clarification of the concept of *coherent superposition* of pure states. As we have hinted above, the Hanle effect is directly related to the generation of coherent superposition of degenerate Zeeman sublevels of an atom (or molecule) by

a light beam.[5] As the Zeeman sublevels are split by the magnetic field, the degeneracy is lifted and the coherences are modified. This gives rise to a characteristic magnetic-field dependence of the linear polarization of the scattered light that is finding increasing application as a diagnostic tool for magnetic fields in astrophysics (e.g., Asensio Ramos, Landi Degl'Innocenti & Trujillo Bueno, 2005).

7. The Paschen-Back effect

As mentioned in Section 2, any atomic level of total angular momentum quantum number J is split by the action of a magnetic field into $(2J+1)$ equally spaced sublevels, the splitting being proportional to the Landé factor g_J and to the magnetic field strength. This well-known result of first-order perturbation theory is correct only if the splitting produced by the magnetic field on a $J-$level is small compared to the energy separation between the different $J-$levels of the (L,S) term under consideration. In other words, the standard theory of the Zeeman effect is valid only in the limit of "weak" magnetic fields. Here, "weak" means that the coupling of either the spin or the orbital angular momentum to the magnetic field is weaker than the coupling between the spin and the orbital angular momentum (the spin-orbit coupling). This is the so-called Zeeman effect regime.

In the opposite limit, the magnetic field is so "strong" that the spin-orbit interaction can be considered as a perturbation compared to the magnetic interaction. In this case the magnetic field dissolves the fine structure coupling – that is, $\vec{L}$ and $\vec{S}$ are practically decoupled and precess independently around $\vec{B}$. Therefore, the quantum number J loses its meaning. In this so-called complete Paschen-Back effect regime the magnetic Hamiltonian is diagonal on the basis $|LSM_LM_S\rangle$, and the term (L,S) splits into a number of components, each of which corresponds to particular values of (M_L+2M_S).

Interestingly, since the spin-orbit coupling increases rapidly with increasing nuclear charge, the conditions for a "strong" field are met at a much lower field with light atoms (like helium) than with heavy atoms. For instance, the levels with $J=2$ and $J=1$ of the upper term 2^3P of the He I 10830 Å multiplet cross for magnetic strengths between 400 G and 1500 G, approximately (e.g., Socas-Navarro, Trujillo Bueno & Landi Degl'Innocenti, 2004). This level-crossing regime corresponds to the incomplete Paschen-Back effect regime, in which the energy eigenvectors are gradually evolving from the form $|LSJM\rangle$ to the form $|LSM_LM_S\rangle$ as the magnetic field increases. This range between the limiting cases of "weak" fields (Zeeman effect regime) and "strong" fields (complete Paschen-Back regime) is more difficult to analyze since it requires a numerical diagonalization of the Hamiltonian.

A particularly interesting phenomenon related to the non-linear effect of the magnetic field on the energy levels in the transition from the Zeeman effect to the complete Paschen-Back effect is the enhancement of the scattering polarization by a vertical magnetic field, which has been investigated by Trujillo Bueno et al. (2002*b*) for the D_2 line of Na I and by Belluzzi, Trujillo Bueno & Landi Degl'Innocenti (2006) for the D_2 line of Ba II.

[5] A *coherent superposition* of two or more sublevels of a degenerate atomic level is a quantum mechanical state given by a linear combination of pure states of the atomic Hamiltonian.

8. Illustrating the effects of the various competing mechanisms on the linear polarization of the He I 10830 Å multiplet

Figure 7 shows theoretical examples of the joint influence of atomic level polarization and the Hanle and Zeeman effects on the emergent Stokes profiles of the He I 10830 Å multiplet, with the wavelength positions and the strengths of the π and σ transitions calculated within the framework of the incomplete Paschen-Back effect theory. This is shown for a broad range of magnetic field strengths, and for both 90° and forward scattering geometries. We point out that the He I 10830 Å multiplet originates between a lower term (2^3S_1) and an upper term ($2^3P_{2,1,0}$). Therefore, it comprises three spectral lines: a 'blue' component at 10829.09 Å (with $J_l = 1$ and $J_u = 0$), and two 'red' components at 10830.25 Å (with $J_u = 1$) and at 10830.34 Å with ($J_u = 2$) which appear blended at solar atmospheric temperatures.

The left panels of Fig. 7 concern the 90° scattering case, which is typical of any off-limb observation. The right panels consider the forward scattering case, which is typical of an on-disk observation at or close to the center of the solar disk. In both cases we have assumed a constant-property slab of optical depth unity at the line center of the red blended component. We have assumed that this slab of helium atoms is located at a height of only 3 arc seconds above the visible solar surface and that it is anisotropically illuminated from below by the photospheric radiation field. The magnetic field is assumed to be horizontal (i.e., parallel to the solar "surface") and oriented in a way such that also for the off-limb case the magnetic field vector is perpendicular to the line of sight. From top to bottom Fig. 7 shows the emergent Stokes Q profile for increasing values of the magnetic strength and for the following three calculations of increasing realism:

(1) The dotted lines indicate the case without atomic level polarization. Here the only mechanism responsible of the emergent linear polarization is the Zeeman splitting of the upper and lower energy levels, which produces wavelength shifts between the π and σ components, whose positions and strengths have been calculated in the incomplete Paschen-Back effect regime. Therefore, zero polarization is found for $B = 0$ G.

(2) The dashed lines correspond to the case in which we have taken into account the influence of the atomic polarization of the two upper levels of the He I 10830 Å multiplet that can carry atomic level polarization, that is, those with $J = 2$ and $J = 1$. Therefore, in addition to the above-mentioned Zeeman effect contribution, we have here the possibility of a selective emission of polarization components, even for the zero field case. For example, this is the reason why the dashed line of the upper left panel of Fig. 7 shows a non-zero linear polarization signal for the off-limb zero field case.

(3) The solid lines correspond to the most general situation in which the influence of the atomic polarization of the lower level is also taken into consideration, in addition to that of the upper levels and to the Zeeman effect. The consideration of lower-level polarization has two consequences. First, the amount of upper level polarization and the ensuing selective emission of polarization components is modified. Second, we can also have a selective absorption of polarization components. For instance, this is the reason why the blue line of the He I 10830 Å multiplet shows a non-zero linear polarization signal in the $B = 100$ G right panel of Fig. 7.

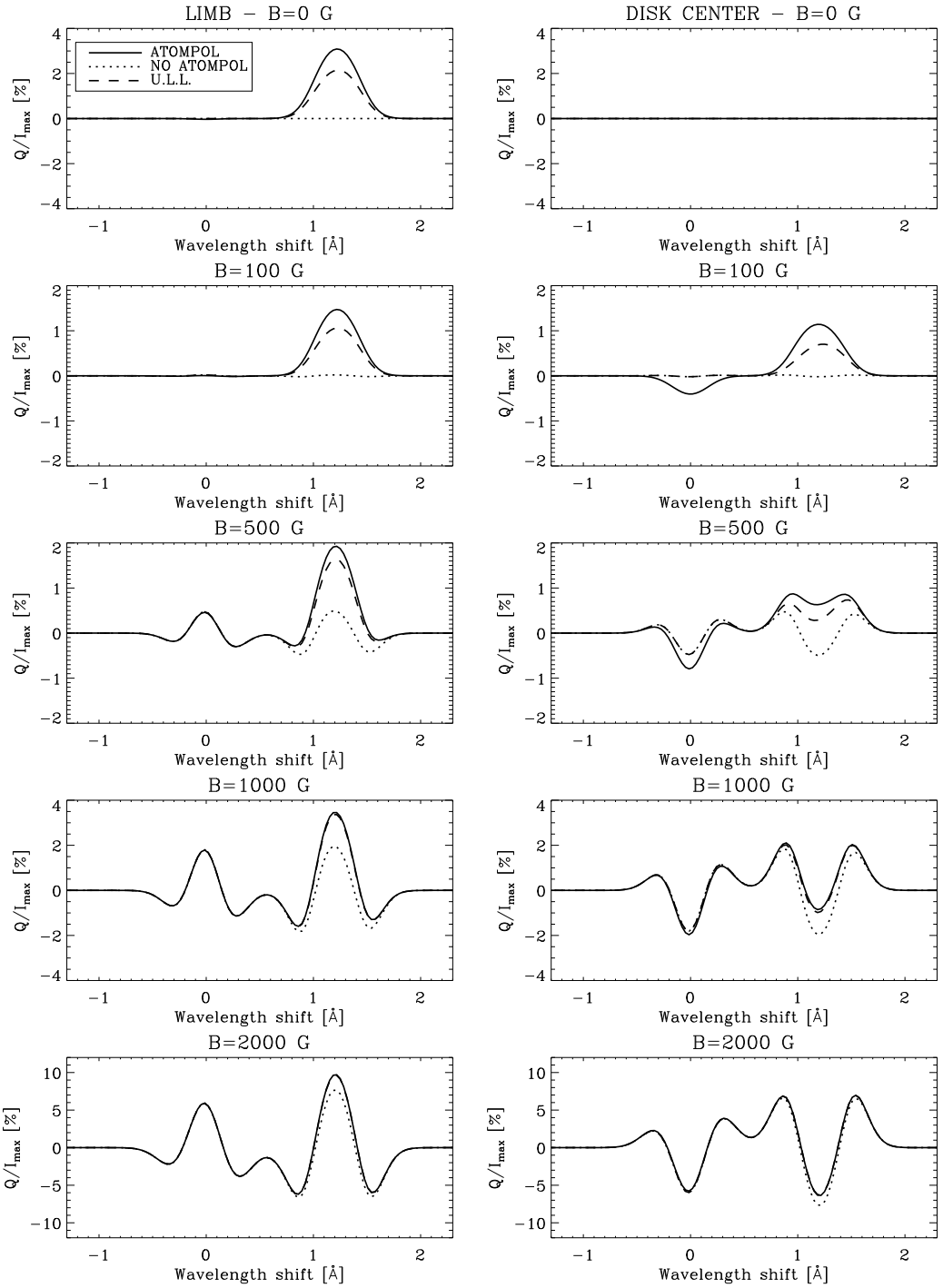


Fig. 7. Influence of atomic level polarization on the emergent linear polarization of the He I 10830 Å multiplet. See the text for details. The positive reference direction for Stokes Q is along the direction of the horizontal magnetic field. From Trujillo Bueno & Asensio Ramos (2007).

As shown in Fig. 7, for weak magnetic fields (e.g., $B \leq 100$ G) the emergent linear polarization is fully dominated by the *selective emission* and *selective absorption* of polarization components that result from the atomic level polarization induced by the anisotropic illumination of the slab's helium atoms (Trujillo Bueno et al., 2002*a*). For stronger magnetic fields, the contribution of the transverse Zeeman effect cannot be neglected. However, the emergent linear polarization may still show an important contribution caused by the presence of atomic level polarization, even for magnetic strengths as large as 1000 G. Interestingly, the observational signature of this atomic level polarization is clearly seen in many of the Stokes Q and U profiles that some observers have measured in emerging magnetic flux regions, even at points of the observed field of view for which magnetic strengths as large as 1000 G are inferred (see details in Trujillo Bueno & Asensio Ramos, 2007). Therefore, the modeling of the Stokes Q and U profiles of the spectral lines of the He I 10830 Å multiplet should *not* be done by taking only into account the contribution of the transverse Zeeman effect, unless the magnetic field intensity of the observed plasma structure is sensibly larger than 1000 G.

9. Summarizing: Zeeman vs. Hanle

The good news about the Zeeman effect is that the mere detection of polarization implies the presence of a magnetic field. The bad news are the following:

- It is of limited practical interest for the determination of magnetic fields in hot coronal plasmas because the Zeeman polarization scales with the ratio between the Zeeman splitting and the Doppler-broadened line width.
- The Zeeman effect is "blind" to magnetic fields that are tangled on scales too small to be resolved.

Concerning the Hanle effect, these are the good news:

- It is sensitive to magnetic fields for which the Zeeman splitting in frequency units is comparable to the inverse lifetime of the upper (or lower) level of the spectral line used, regardless of how large the line width due to Doppler broadening is. It is therefore sensitive to weaker magnetic fields than the Zeeman effect: from at least milligauss to hundreds of gauss.
- It is sensitive to the presence of hidden, mixed-polarity fields at sub-resolution scales.
- Contrary to a widespread belief, the diagnostic use of the Hanle effect is *not* limited to a narrow solar limb zone. In particular, in forward scattering at disk center, the Hanle effect can create linear polarization, when in the presence of inclined magnetic fields.

The downside of the Hanle effect is that it is properly a quantum effect, and the quantum theory of polarization is a complicated subject. However, it has been recently described in great detail in a rigorous monograph (see Landi Degl'Innocenti & Landolfi, 2004). And we know how to solve the relevant equations accurately and efficiently in order to model polarization phenomena in (magnetized) astrophysical plasmas (Trujillo Bueno, 2003*b*; see also the multilevel radiative transfer code of Manso Sainz & Trujillo Bueno, 2003*b*).

10. Concluding remarks

Polarized light provides the most reliable source of information at our disposal for the remote sensing of astrophysical magnetic fields, including those on the Sun. However, in order to open a true empirical window on solar and stellar magnetism, we need to develop and apply suitable diagnostic tools that take proper account of the various physical mechanisms that produce polarization in spectral lines. Hopefully, the brief introduction provided here will motivate the reader to go deeper into the subject.

Acknowledgements. This research has been funded by the Spanish Ministerio de Educación y Ciencia through project AYA2004-05792.

References

Asensio Ramos, A., Landi Degl'Innocenti, E., Trujillo Bueno, J. 2005, ApJ, **625**, 985
Belluzzi, L., Trujillo Bueno, J., Landi Degl'Innocenti E, 2006, in Solar Polarization 4, eds. R. Casini & B. Lites, ASP Conf. Series, in press
Born, M., & Wolf, E. 1994, *Principles of Optics*, Pergamon Press, Oxford
Condon, E.U., & Shortley, G.H. 1935, *The Theory of Atomic Spectra*, Cambridge University Press, Cambridge
Gandorfer, A. 2000, *The Second Solar Spectrum*, Vol. 1: 4625 Å to 6995 Å ISBN 3 7281 2764 7 (Zürich: vdf)
Gandorfer, A. 2002, *The Second Solar Spectrum*, Vol. 2: 3910 Å to 4630 Å ISBN 3 7281 2855 4 (Zürich: vdf)
Hanle, W. 1924, Z. Phys., **30**, 93
Happer, W. 1972, Rev. Mod. Phys., **44**, 169
Kastler, A. 1950, J. de Physique, **11**, 255
Landi Degl'Innocenti, E., Landolfi, M. 2004, *Polarization in Spectral Lines*, Kluwer Academic Publishers
Manso Sainz, R., Trujillo Bueno, J. 2003*a*, Phys. Rev. Letters, **91**, 111102-1
Manso Sainz, R., Trujillo Bueno, J. 2003*b*, in Solar Polarization 3, eds. J. Trujillo Bueno & J. Sánchez Almeida, ASP Conf. Series **307**, 251
Moruzzi, G., Strumia, F. (eds.) 1991, *The Hanle Effect and Level-Crossing Spectroscopy*, New York: Plenum
Socas-Navarro, H., Trujillo Bueno, J., & Landi Degl'Innocenti, E. 2004, ApJ, **612**, 1175
Stenflo, J.O. 1982, Solar Phys., **80**, 209
Stenflo, J.O. 2002, in *Astrophysical Spectropolarimetry*, J. Trujillo Bueno, F. Moreno-Insertis & F. Sánchez (eds.), Cambridge University Press, 55
Stenflo, J.O., & Keller, C. 1997, A&A, **321**, 927
Stenflo, J.O., Keller, C., & Gandorfer, A. 2000, A&A, **355**, 789
Trujillo Bueno, J. 1999, in Solar Polarization, eds. K. N. Nagendra and J. O. Stenflo, Kluwer Academic Publishers, 73
Trujillo Bueno, J. 2001, in Advanced Solar Polarimetry: Theory, Observation and Instrumentation, ed. M. Sigwarth, ASP Conf. Series **236**, 161
Trujillo Bueno, J. 2003*a*, in Solar Polarization 3, eds. J. Trujillo Bueno & J. Sánchez Almeida, ASP Conf. Series **307**, 407
Trujillo Bueno, J. 2003*b*, in Stellar Atmosphere Modeling, eds. I. Hubeny, D. Mihalas and K. Werner, ASP Conf. Series **288**, 551
Trujillo Bueno, J., Landi Degl'Innocenti, E. 1997, ApJ Letters, **482**, L183
Trujillo Bueno, J., Asensio Ramos, A. 2007, ApJ, **655**, 642
Trujillo Bueno, J., Landi Degl'Innocenti, E., Collados, M., Merenda, L. & Manso Sainz, R. 2002*a*, Nature, **415**, 403
Trujillo Bueno, J., Casini, R., Landolfi, M., Landi Degl'Innocenti, E. 2002*b*, ApJ Letters, **566**, L53
Trujillo Bueno, J., Shchukina, N., Asensio Ramos, A. 2004, Nature, **430**, 326

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

SOLAR MAXIMUM STREAMERS AS THIN TWISTING SHEETS

H. MORGAN and S. RIFAI HABBAL
Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

Abstract. Images of the corona processed using new techniques reveal detailed fine-scale structure and provide accurate depictions of large-scale structure. Solar maximum streamers which appear at high latitude in white light images are shown to appear and disappear with solar rotation. Carrington-type maps reveal the thin, highly filamentary sheet-like structures of these streamers. Examples of thin sheet structures at very low heights in the corona can also be seen in EIT images, which suggests that some solar maximum streamers exist as twisting sheets from the chromosphere out to the outer corona. This proposition is explored using a structural model of the corona, and the complex appearance of a streamer is well simulated using a twisting sheet topology.

Key words: solar physics, corona

1. Normalizing Radial Graded Filter

The normalizing radial-graded filter (NRGF), and other image processing techniques, are described in detail by Morgan *et al.* (2006). The NRGF is a simple and accurate method to look at the large-scale structure of the corona. It calculates the average and standard deviation of brightness as a function of height within an image, and uses these values to normalize the brightness at each height, therefore viewing the coronal structure seen in a NRGF-processed image is equivalent to comparing a large set of normalized latitudinal profiles simultaneously.

The results of applying the NRGF to images of the solar minimum and maximum corona are shown in Figure 1. These images are composed of observations in He II 304 Å by the Extreme Ultraviolet Imaging Telescope (EIT) (Delaboudiniere *et al.*, 1995), and pB observations by the Mauna Loa Solar Observatory's (MLSO) MKIII (solar minimum) and MKIV (solar maximum) coronameters (Fisher *et al.*, 1981) and the LASCO C2 coronagraph (Brueckner *et al.*, 1995). As can be seen, a useful aspect of the NRGF is the ease it allows to join images from various instruments. The simple normalizing function results in excellent continuity at the boundaries between the instruments'field of views, without interpolation or smoothing. This suggests the correctness of using the NRGF to look at structure.

2. Streamers as Thin Sheets

Figure 2 shows two high-latitude streamers in the north corona labeled A and B. They appear and disappear with a $\sim$7 day interval. When the streamers are seen (left column), the region between the streamers above the North pole is darkest. When the streamers cannot be seen (right column), the same region is brighter and contains some structure. So in the field of view of LASCO C2, streamers A and B are likely

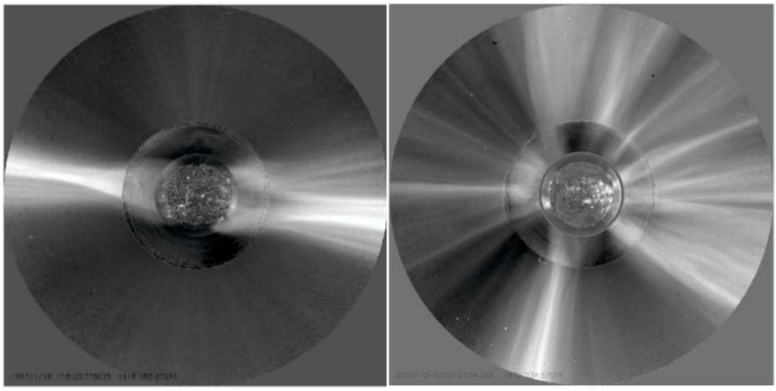


Fig. 1. Composite coronal images processed by the NRGF for dates 1997/01/18 (left) and 2000/12/03 (right).

to be thin sheet-like structures. In the left column, the streamers appear narrow due to being viewed edge-on, and are bright due to their extended contribution along the line of sight. Seven days later the Sun has rotated through a quarter turn and the streamers are viewed face-on above the pole. The bright narrow structures have disappeared and above the North pole their broad and diffuse filamentary structure is revealed. The sheet-like density structure of streamers at heights above $\sim 2.5 R_{\odot}$ is explored in detail by Thernisien and Howard (2006).

Between 2000/12/02 19:31 and 2001/01/12 23:54 the LASCO C2 coronagraph made over 2400 total brightness observations of the corona. These images are calibrated using the standard LASCO procedures included in the Solar Software package, and processed using procedures described in Morgan *et al.* (2006). From these images we extract horizontal slices of processed brightness at various heights above the North pole. For illustration, the position of two slices are shown as the dotted horizontal lines in the right bottom image of Figure 2. The slices are stacked in time to create maps such as the one shown in Figure 3. Within the map, bright structures at high latitude describe sinusoidal curves with solar rotation. The periodic appearance of the high-latitude streamers labeled A and B in Figure 2 is clearly shown in Figure 3 as an effect of rotation. The approximate alignment of these structures along the line of sight result in the brightenings labeled A and B. The periods of alignment, or brightening can be seen to last for 2 to 4 days. From these synoptic maps, and the face-on appearance of the streamers in LASCO C2 images and movies, we can say that the streamers are thin, high-density filamentary sheets.

The source of streamer B is a narrow but long region of considerable activity seen in EIT 171Å observations, shown in the left image of Figure 4. When this region is centered over the east limb, a bright thin sheet of material can be seen emerging directly from the disk to the corona. Observing the passage of the source region against the limb over the course of many days reveals direct evidence of an extended thin sheet, as shown in the right image. The sheet is also apparent half a rotation later on the west limb. We have established in this section that streamers A and B exist as thin sheets in the corona. EIT observations show that these streamers may

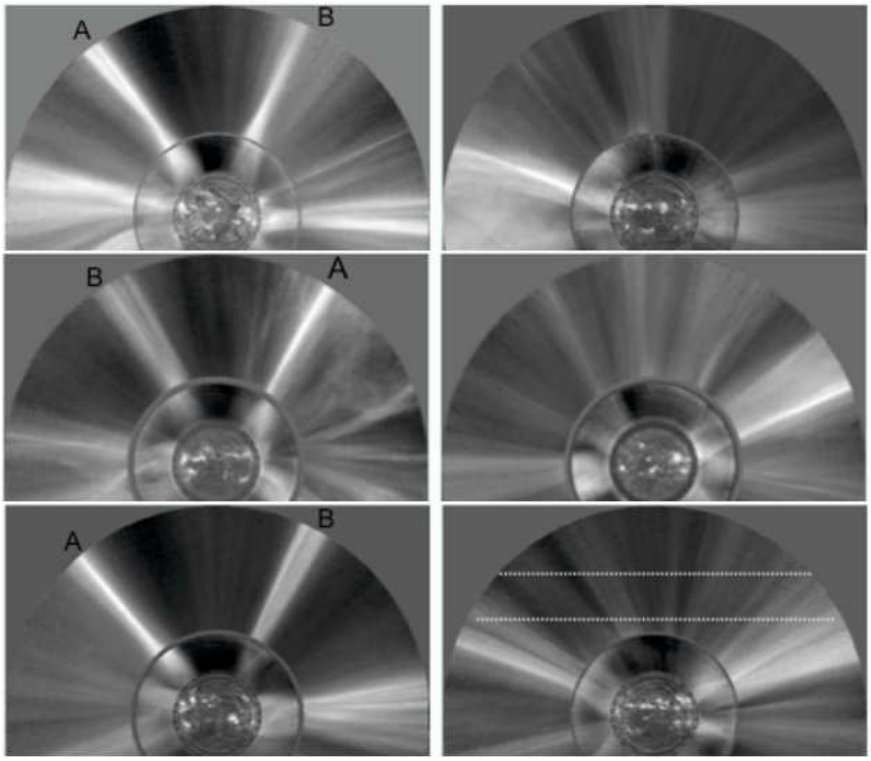


Fig. 2. Composite images of the Sun and corona made during Carrington rotations 1970 and 1971 at approximately 7 day (quarter rotation) intervals. Two high latitude streamers are labeled A and B.

exist as thin sheets at very low heights.

3. Exploring the Structure of Streamers

To explore the structure of coronal streamers from the solar surface to heights above where the corona becomes radial, a density model is constructed which allows the flexibility to incorporate complex structures. The position of high-density structures near the solar surface can be defined by, for example, the position of prominences as observed in Hα, or by bright regions in EIT 171Å images. The position of structures in their final radial configuration at $\sim 3R_{\odot}$ can be adjusted until images produced from the model agree with LASCO C2 observations. The thickness and position of structures evolve smoothly from the solar surface to $3R_{\odot}$, then maintain a radial configuration. Synthetic pB images are created from the coronal density model and are compared to observations. No attempt is made to determine the absolute electron density for the background corona or streamers. The background density is zero, and the density of streamers is set at an arbitrary value of 1 at the solar surface, decreasing with the square of the height. The brightness images obtained by line of sight (LOS) integration of the density model are therefore in arbitrary

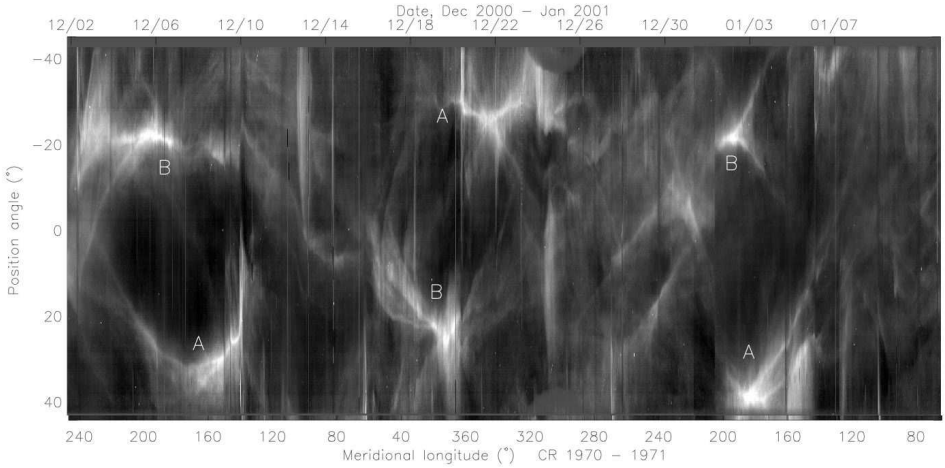


Fig. 3. LASCO C2 synoptic map, showing slices at a central height of $4.5R_{\odot}$. The y axis is position angle (positive is degrees counter-clockwise from North). Periodic regions of higher brightness are labeled A and B, corresponding to the labeled streamers in Figure 2.

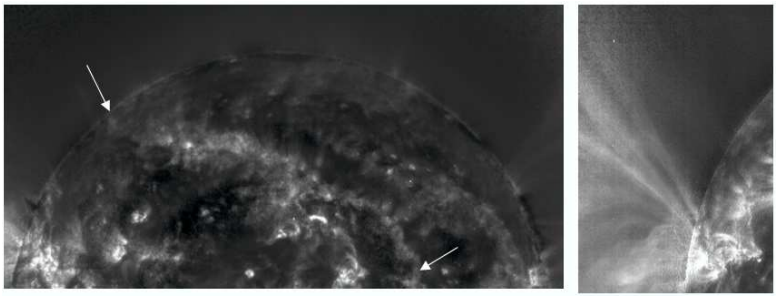


Fig. 4. *Left* - Source region of streamer B seen as an extended narrow region of enhanced activity in an EIT 171Å image. Arrows point to both ends of the region. *Right* - Example of a narrow thin sheet emerging directly from the region shown in the left image.

units, and are zero for any LOS not containing a streamer. The synthetic images are 'flattened' using the NRGF, and compared to observations, also processed with the NRGF. This is a valid and computationally efficient method to explore the large-scale coronal structure.

From movies made from EIT observations, we can see much dynamic activity in the source region of streamer B (shown in Figure 4), and we believe this causes large changes in the streamer. This makes streamer B very difficult to model. We concentrate our efforts therefore on streamer A. The top plot of Figure 5 shows the source of streamer A as a long filament (or prominence) as observed in Hα, with some active regions associated with the filament at lower latitudes (larger longitudes). This filament lies along a magnetic neutral line in the photosphere, as shown in the photospheric field synoptic map of the bottom plot. We model streamer A as a narrow high-density structure arising directly from the filament. The position of

streamer A at the model solar surface is shown as the red region in Figure 5. We adjust the position of streamer A in the higher corona (the model corona becomes radial at $\sim 3R_\odot$) until modeled coronal images agree with observations made over the course of half a solar rotation. This position is shown as a yellow region in Figure 5. It is interesting that this coronal configuration has no agreement with the position of the coronal neutral line as calculated by source-surface extrapolation of the photospheric field.

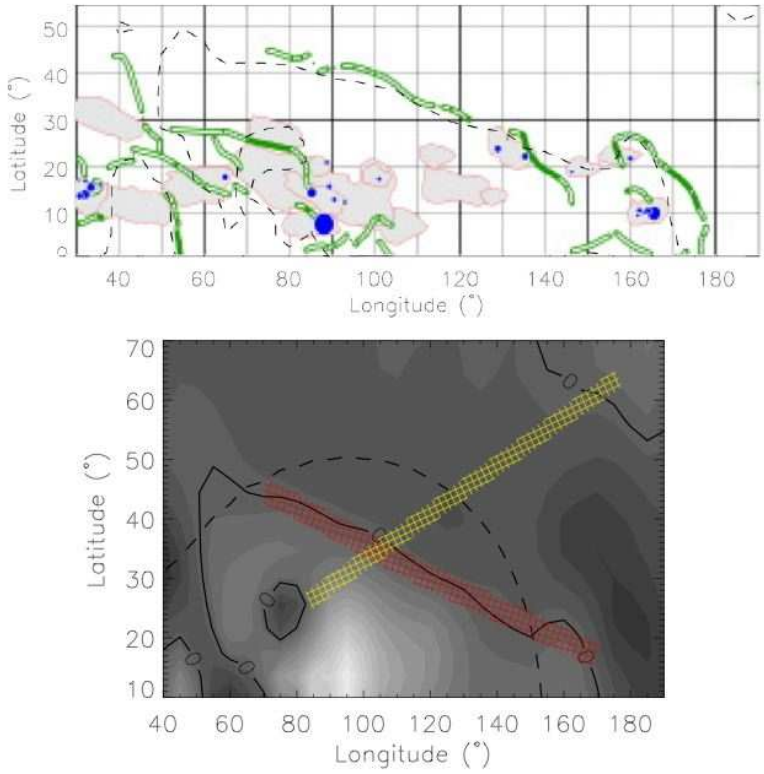


Fig. 5. *Top* - Source region of streamer A as seen in a Paris-Meudon Observatory (PMO) synoptic map. Features shown are filaments observed in Hα (green lines), sunspots as seen in a Ca II line (blue spots) and plage observed in a Ca II line (grey areas). The source of streamer A is the long filament (or prominence) lying from longitude 70°, latitude 50° to longitude 170°, latitude 10°. Also associated with this filament are active regions (sunspots and plage) at lower latitudes. The x-axis gives Carrington longitude for rotation 1970. *Bottom* - Modeled configuration of streamer A overlaid on a synoptic map of the photospheric magnetic field measured by the Wilcox Solar Observatory (WSO). The red and yellow regions show the modeled configuration of streamer A on the solar surface and in the high corona respectively. Brighter regions in the photospheric field map show a stronger positive field. Neutral lines are shown as solid bold contours. The dashed bold contour shows the neutral line at a height of $3R_\odot$, as calculated by WSO using a source-surface extrapolation of the photospheric field.

Figure 6 shows that considerable success is achieved in simulating streamer A as a twisting thin sheet. The changing apparent latitude of the main stalk is respected.

The changing shape of the streamer with rotation is also well replicated. The bottom two pairs of images, for rotation 150° and 124°, show replication of a faint structure at one limb (labeled c and d) simultaneously with the main body of the streamer at the other limb - structure which can not easily be associated with streamer A without the support of the model. Note also the appearance of the streamer as a fan-shaped structure in the top image. This is due to the edge-on alignment of the sheet at low heights in the corona. The twisting of the sheet with increasing height gives a diverging, fluted shape to the streamer since it is seen more face-on higher in the corona. The opposite is true in the second image (for longitude 176°). The sheet is seen more face-on low in the corona, and narrows to edge-on with increasing height. This gives a helmet-shaped streamer.

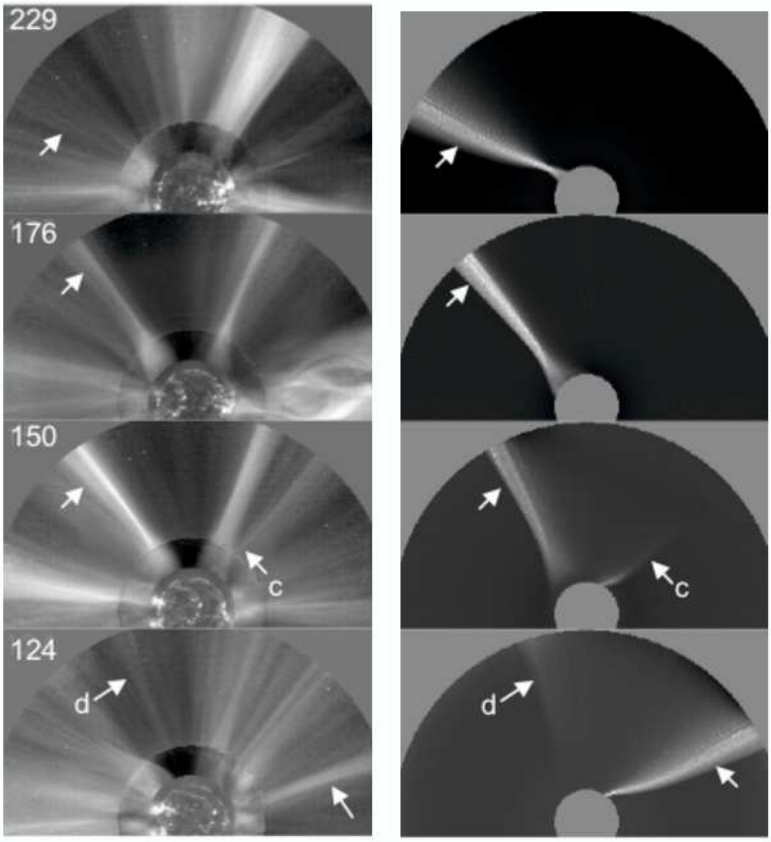


Fig. 6. Four comparisons of the observed north corona (left) and the appearance of streamer A modeled as a twisting sheet (right column). The numbers in the top left give the Carrington longitude . From top to bottom, the observations correspond to dates 2000/12/03, 07, 09 and 11 respectively. Arrows help identify features formed by streamer A. Two faint features are labeled c and d.

4. Summary

The comparison of synthesized and observed coronal images processed by the NRGF is a simple and efficient method to explore the structure of the corona without concern for the absolute coronal density, or density as a function of height. It allows the quick exploration of structure and the testing of ideas against observation. It can also be used as a basis for calculating absolute densities, since once the structure is 'fixed' in the model, and a background density found for a streamerless region of the corona, the density of structures may readily be found by inversion.

The main points of this work are:

- LASCO C2 images and synoptic maps show the thin sheet topology of two high-latitude solar maximum streamers.
- Direct observations of thin sheets at low heights are often seen in solar maximum EIT images. These sheets are not part of large closed-field structures, since they are seen to extend into the corona and form the base of streamers.
- Simulating streamers as thin twisting sheets closely replicates the appearance of complex structures in the solar maximum corona, and their changing appearance with solar rotation.
- We show that twisting sheets can give the appearance of helmet streamers. At other alignments, they can appear as fan streamers. So a simple topology gives an elegant interpretation for two categories of streamers with a very different appearance.
- In successfully simulating streamers in the solar maximum corona, we find no agreement between the position of the modeled streamers in the corona and the position of the coronal neutral line calculated by source-surface extrapolation of the photospheric field.

Acknowledgements. SOHO is a mission of international cooperation between ESA and NASA. The Wilcox Solar Observatory (WSO) data and model results were obtained from the WSO section of Stanford University's website courtesy of J.T. Hoeksema. WSO is supported by NASA, the NSF and ONR. The PMO maps are available from *http://bass2000.obspm.fr/* and the WSO data from *http://soi.stanford.edu/wso/*.

References

Brueckner, G. E., et al.: 1995, *Solar Phys.*, **162**, 357
Delaboudiniere, J.-P., et al. 1995, *Solar Phys.*, **162**, 291
Fisher, R. R., Lee, R. H., MacQueen, R. M., and Poland, A. I., 1981, *Applied Optics*, **20**, 1094
Morgan, H., Habbal, S.R., & Woo, R. 2006, *Sol. Phys.*, *in press* (available at http://xxx.lanl.gov/abs/astro-ph/0602174)
Thernisien, A. F., & Howard, R. A. 2006, *ApJ*, **642**, 523


olar Physics a
lipses
Libysche - Deutsche

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

A PROPOSAL FOR SOLAR GRAVITATIONAL REDSHIFT MEASUREMENT

ALESSANDRO CACCIANI
Laboratorio di Fisica Solare Università La Sapienza Roma

RUNA BRIGUGLIO
Laboratorio di Fisica Solare Università La Sapienza Roma

FABRIZIO MASSA
INFN sez Roma

and

PAOLO RAPEX
External collaborator

Abstract. The Gravitational Red Shift (**GRS**) is the loss of energy of a photon ascending a gravitational potential. Even if it can be interpreted in terms of classical mechanics, deviations from its theoretical value imply considerations about General Relativity. Hence, a precise determination of its value will be an independent test of the theory. Because of its large mass, compared to Earth, the Sun is a preferred source to be considered to study this effect; nevertheless, up to now, the best precision of the solar GRS is quoted not better than 1%, as from the historical measurement of Lo Presto. A new acquisition hardware has been developed at the Solar Laboratory of the University of Rome. It takes advantage of a Magneto Optical Filter (**MOF**), a very narrow bandpass filter, tuned on Na or K lines, that allows to acquire Doppler images of the sun. The starting point of this technique is that sun shows a very large Doppler signal due to its rotation: the only points displaying zero values are those close to the rotation axis. Their separation is indeed a measure of the solar GRS. With our instrumentation, the GRS can be measured with high precision as some preliminary analysis on previous data demonstrate. This presentation at SPSE meeting is for large part taken from a paper of the same authors accepted by the journal Celestial Mechanics and Dynamical Astronomy.

Key words: solar physics, GRS, MOF, solar lines, Doppler signal, general relativity

1. Introduction

The gravitational potential of a celestial body affects the physical time, slowing down any periodic phenomenon assumed as the physical clock. In order to verify this statement, the behavior of different types of clocks, located in different gravitational potentials, must be compared. During the last decades, four main experiment attempted to provide a measurement of gravitational redshift. The first is the famous Pound-Rebka-Snider experiment (Pound *et al.*, 1960) using the Mossbauer effect and measuring the frequency shift of γ-rays ascending a 22.6 m. tower at Harvard. The accuracy was of the order of 10%. In 1980 Vessot compared the frequency standard of two Hydrogen Masers, one on the ground and the other at 10000 Km on a spacecraft, was able to verify the theoretical prediction at the level of about $2 \cdot 10^{-4}$. In 1993 a similar experiment was conducted on board Galileo spacecraft

by Krisher, using radio signals, and achieving a precision of 1%.

A different kind of possible experiments is provided by the solar spectral lines. The solar gravitational potential shifts their wavelengths relative to laboratory lines on Earth by the amount of $\approx 2.1 \cdot 10^{-6}$ towards the red. In terms of velocity, this is equivalent to a Doppler shift of 636 m/s outside the Earths gravitational field and 633 m/s at the Earths surface. Lo Presto ((LoPresto *et al.*, 1991)) used chromospheric Oxygen lines in emission, that are formed well above the main convective velocity field, reaching the precision of about 2%. The solar spectrum is attractive in this context because the Sun is a massive body monitoring the GRS to values far larger than what is possible on Earth (dynamic range).

2. Solar measurement

A number of problems arise when attempting to measure the GRS using the solar spectrum. One has to consider that, while helioseismology deals with differential signals (so that the need for accurate bias or systematic errors removal is not so strong) in the GRS case the wanted signal is itself a constant offset. Hence, one has to take into account a number of bias contributions, partly instrumental, partly arising from the solar lines themselves. As an example, the entire visible solar surface seems to move towards the observer as a consequence of convective motions; the result is a global, non uniform (turbulent, convective) blueshift. Helioseismology, on the other hand, brings by itself another source of noise: oscillating Doppler signals are superimposed on the GRS signature.
Let's consider the resulting equation from all the sources of Doppler shifts:

$$V_{Obs} = V_{GRS} + V_{BS} + V_{LE} + V_{dR} + V_{EOS} + V_{IB} + V_{He} \tag{1}$$

where V_{Obs} is the overall observed Doppler signal from a point on the Earth's surface (the observatory),V_{GRS} is the amount of the solar gravitational redshift relative to the same point, V_{BS} is the convective blue-shift, V_{LE} is the limb effect, V_{dR} is the contribution of solar differential rotation, V_{EOS} accounts for terrestrial rotation and sun-earth motion, V_{IB} takes into account instrumental biases and V_{He} is the helioseismic signal.

Among all these sources of noise, only V_{EOS} can be easily removed (from the ephemeris); for the other effects one can only rely on suitable models. As an example, convective blue shift exhibits a cosine dependence from the disk center (minimum at the solar limb, maximum on its center). Viceversa the limb effect is maximum at limb, minimum at disk center. Similar considerations (see also section 4) lead to the conclusion that an imaging Doppler acquisition is a necessary and efficient method to deal with this kind of noises. The instrument we use to acquire solar Doppler images has been developed during last decades in Rome by Cacciani ((Cacciani *et al.*, 1990),(Cacciani *et al.*, 2001)) and is called the Magneto Optical Filter (MOF).

3. The Magneto Optical Filter

The MOF is the core of the system that makes this experiment unique. Here let be enough to say that it is a very stable and narrow double band filter. The two bands

could be as narrow as $50m\mathring{A}$, achieving unsurpassed performances as far as its central wavelength reference stability and symmetrical tuning in the wings of the solar lines (Red and Blue sides). Each band can be selected separately at will so that a computer comparison (difference) between the transmitted images can produce the wanted Doppler image, while their sum gives an intensity image. As a filter, the MOF will be located between the telescope and the image sensor. In this manner we are able to reject all the other wavelengths of the solar spectrum, but the wanted line, so that we can definitely say that the MOF produces an artificial night, which is the necessary condition to detect faint signals. As an example, in Figure 1 (which is an image tuned in the core of the Sodium D lines) an intensity reversal (faint emission) is visible in the narrow MOF band-pass wherever magnetic field emerges from below at photospheric levels. Indeed, the magnetic field excites the emitting atoms and modifies the line profiles: the consequent slight increase of its central luminosity appears well visible in our MOF image. This effect should be taken in due consideration to obtain a precise determination of the line shift because it creates macroscopic distortions of the velocity measurement: magnetic regions should therefore be avoided, excluding them from the analysis. In Figure 1 the predominant rotation signal (from the East limb to the West limb) amounts to about 4 Km/sec. The signal originated by the GRS amounts to 633 m/sec (Doppler equivalent). We have been able to measure Doppler signals as low as 1 mm/sec by integration over the whole solar disk. Figure 2 shows a plot of the oscillatory signal due to solar p-modes in the 5 minute band (peak-to peak amplitudes 1 m/sec , integration time=30 seconds, telescope aperture 2 cm; JPL facility, Pasadena, California); therefore, our instrument has the potential capability to improve considerably the accuracy beyond the few percent so far achieved. Our final goal is to reach the precision of at least one part per mil.

The MOF's weight (about 1 Kg) and dimensions (10x10x30 cm^3) are very attractive for space applications. In this context, we could imagine a space project aimed to test the second order gravitational time dilation (a real test of the General Relativity metrics) going close to the Sun. A comparison between the solar signals from two satellites, one near the Earth and the other close to the Sun, will cancel all the unwanted effects listed above, but the GRS alone. In fact GRS depends on distance Sun-telescope, while the sources of noise are common in the two cases. The second order effect is proportional to ΔU^2 (where U is the gravitational potential and δU is its difference between the source and the detector) and is in the range of a mm/s (Doppler equivalent), well within the capability of our MOF. A space mission is also desirable to avoid spurious effects from the Earth's atmosphere; however, due to the high costs and other difficulties of space projects (as, for example, the spacecraft motions), it is preferable to perform our first measurements from the ground, taking advantage of the absolute wavelength reference of the MOF and the well known Sun-Earth relative Doppler shifts at any time.

4. First results

For a full and complete description of preliminary analysis and results, please refer to (Cacciani *et al.*, 2006). Now, just few comments about the advantages of using

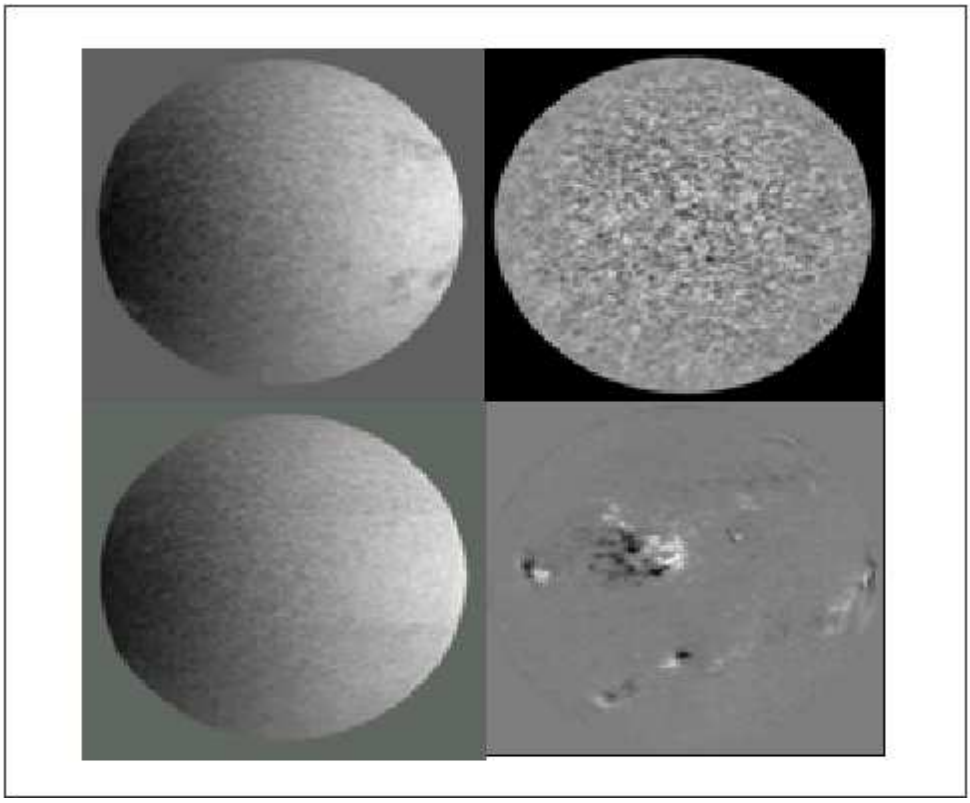


Fig. 1. Image of sun disk acquired with MOF. The left panel show the Doppler effect due to sun rotation. The right panel shows the Doppler signal when rotation is removed (upper) and the magnetogram of sun disk (lower)

full-disk Doppler images:

- It is possible to identify and avoid the magnetic regions during the analysis (see the dark patches in Figure 1.
- Each pixel can be normalized to the local intensity (that is not homogeneous over the solar disk). This could prevent any crosstalk between Doppler and Intensity in the case we have very high image resolution, practically unachievable. Indeed, included in each resolution element coexist microscopic rising and falling elements, the first being hotter and brighter than the descending ones, so that a net blue shift results when the image resolution is poor.
- It is possible, pixel by pixel, to account for the fact that the above blue shift, V_{BS} is maximum at the solar centre and vanishes at the limb as $\cos\phi$.
- The solar rotation signal, V_{dR}, although differential in latitude, appears always null at the axis where it is perpendicular to the line of sight.
- Finally, although the effect of V_{LE} (the centre to limb variation of the solar line profile) is more difficult to be accounted for, it is possible to evaluate its contribution assuming a linear behaviour with the distance from the disk centre,

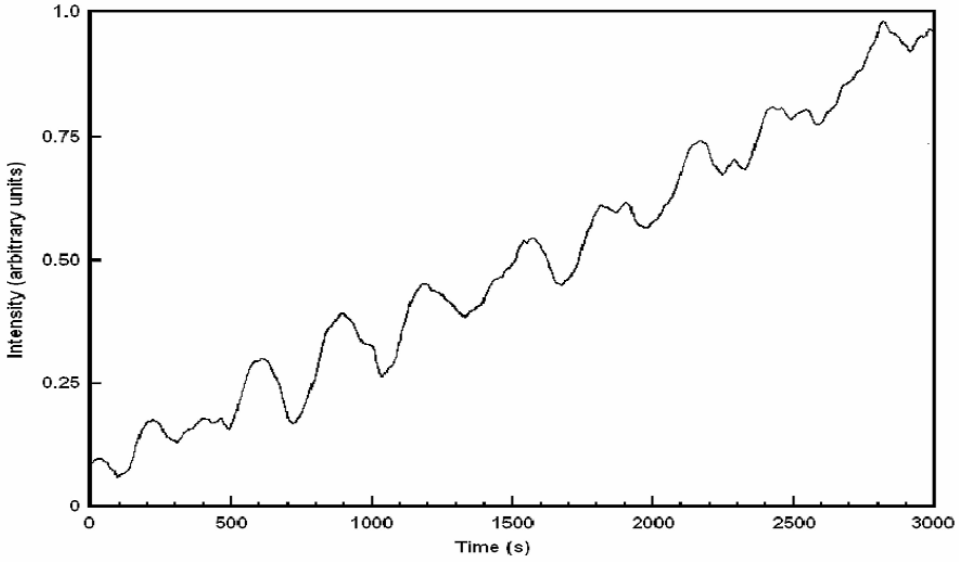


Fig. 2. A plot of helioseismology signal with MOF: peak to peak amplitude is of the order of 1 m/s, superimposed to the Earth rotation effect.

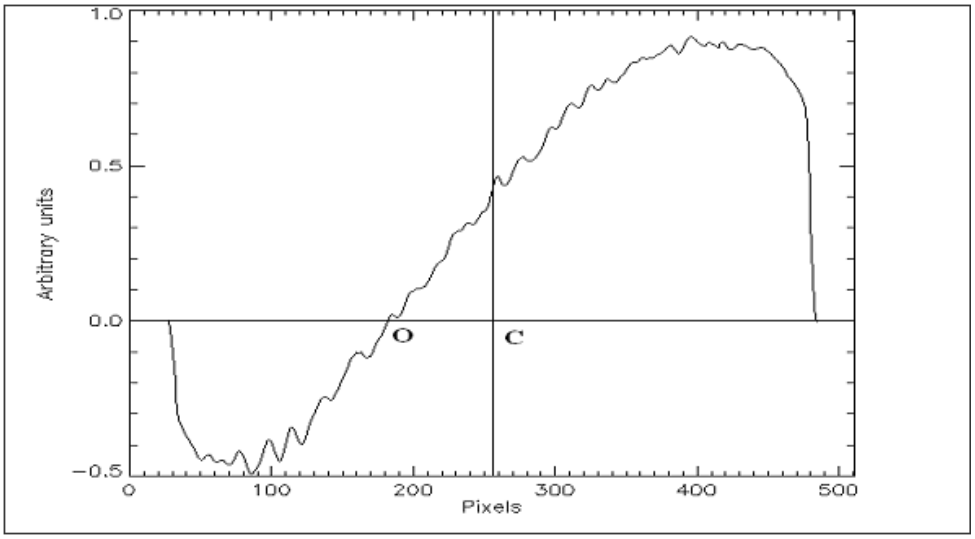


Fig. 3. Plot of Doppler effect signal on sun equator: the zero-crossing point displacement from rotation axis is the indicator of solar GRS.

within a limited region (inside the larger circle in Figure 1.

Each Doppler image will thus offer many points where the full signal can be computed (redundant system of equations). The Ephemeris data will help determining the values for V_{dR} and V_{EOS} to a very high degree of accuracy, while the two quantities

V_{BS} and V_{LS} will be obtained measuring distances in pixels (as both are assumed to be linear functions of distance, inside the limited central region). With a large time series of Doppler images we are confident to achieve our result with an error within the 10^{-3} limit. Indeed, our software procedure to identify image centres and diameters has shown that the time series of those quantities is normally distributed, allowing controlling the degree of accuracy as a function of length of the time series.

4.1. Data analysis

The full project is quite expensive and several funding proposals have been submitted. While waiting for the necessary financial support, we have analyzed a few MOF Doppler images, including some taken previously during our 1999 expedition at the Antarctic Italian site of Baia Terranova. It is remarkable the fact that just looking at the Figure 1, the zero velocity line appears crossing the equator at about one quarter radius eastward which corresponds to about 500 m/s taking into account the line of sight component of the solar rotation only.

We shall limit the analysis within a given circle around the centre, such as that the two signals V_{BS} (which is maximum at the centre) and V_{LE} (which is null at the centre and maximum at the limb) vary linearly with the distance r from C:

$$V_{BS} = (1 - r)V_{BS}(C)$$

$$V_{LE} = r \cdot V_{LEmax}.$$

For the centre C and all the points L_i along the rotation axis, $V_{dR} = 0$ and the equation 1 becomes

$$V_{Obs} = V_{GRS} + V_{BS}(C) + (V_{LEmax} - V_{BS}(C)) \cdot r + V_{EOS} + V_{IB} + V_{He} =$$

$$V_{GRS} + V_{BS}(C) + \gamma \cdot r + V_{EOS} + V_{IB} + V_{He}$$

where $\gamma = (V_{LEmax} - V_{BS}(C))$. Using this equation for the centre C and another point L_i , we can determine the value of γ:

$$V_{Obs}(Li) = -\gamma \cdot r_{Li} + \Delta V_{He}$$

$$\gamma = V_{Obs}(Li) - V_{Obs}(C))/r_{Li} + \Delta V_{He}.$$

The term ΔV_{He} can be made negligible just smoothing the signal along the rotation axis and taking its average from a long time series of images. For a rough signal calibration, we can use eqn a) for a point L_i on the axis and a point P off axis, but at the same distance from C. We get

$$V_{Obs}(P) - V_{Obs}(Li) = V_{dR}(P),$$

which is well known. The final results of such calibration leads $V_{GRS} = 625 m/s$. The above result is already at the level of 1% internal error; however, it is only indicative being affected by errors, unknown biases and long term instabilities that our project will be able to characterize very precisely.

A better calibration will come from the orbital motion of the Earth during a full year of data.

We thank James LoPresto, Jacques Beckers for their comments and the Observatories of San Fernando (California) and Locarno (Switzerland) for their availability to host our instrumentation for the future accomplishment of the project. We thank the Italian PROGETTO ANTARTIDE for the use of Baia Terranova site and financial support.

References

Cacciani, A., Rosati, P., Ricci, D., Egidi, A., Apporchaux, T., Marquedant, R., Smith, E.: 1990, JPL Document D11900 Theoretical and experimental study of the Magneto-Optical Filter

Cacciani, A., Moretti, P.F.: 1993, in Astronomical Telescopes and Instrumentation for the 21st Century, SPIE, **2198**, 219

Cacciani, A., Dolci, M., Moretti, P.F., D'Alessio, F., Giuliani, C., Micolucci, E., Di Cianno, A.: 2001, A&A, **372**, 317

Cacciani, A., Briguglio, R., Massa, F., Rapex, P.: 2006, Celestial Mechanics and Dynamical Astronomy, **95**, 425

Dravins: 1982, ARA&A, **20**, 61

LoPresto, Shrader and Pierce: 1991, ApJ, **367**, 757

Pound, Rebka, Snider: 1960, Phys. Rev. Lett., **4**, 337

Stix, M.: 1972, Phys.Rev.Lett. **28**, 853

Stix, M.: 2004, The Sun, Springer II Ed. 2004, 262, fig.6.13


232-E
231-F

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

MEASUREMENTS OF ATOMIC AND MOLECULAR PARAMETERS OF HYDROGEN AND NITROGEN FOR SOLAR PHYSICS

ADRIAN DAW[*], ANTHONY G. CALAMAI[*], SAMUEL BREWER[†] and BRIAN MYER[‡]
Department of Physics and Astronomy, Appalachian State University, Boone, NC 28608, USA

and

DANIEL WOLF SAVIN and MICHAEL SCHNELL
Columbia Astrophysics Laboratory, Columbia University, New York, NY 10027, USA

Abstract. Experiments in progress at Appalachian State University's Ion Trap Laboratory are providing atomic and molecular data for solar, stellar, planetary, and astrophysical plasmas: collision rates coefficients, radiative decay rates of metastable ions, and unimolecular dissociation rates of ionized molecules. Processes currently under study include: collision rates of protons with H_2 and He, of atomic and molecular nitrogen ions with N_2, radiative decay of metastable $2s2p^3\ {}^5S_2$ N^+, and the dissociation rate of doubly-ionized molecular nitrogen. The first results of these investigations are presented, and the radio-frequency ion trap apparatus and its many capabilities to provide data for solar physics and related fields are discussed.

Key words: atomic and molecular data, ultraviolet spectroscopy, solar activity, sunspots, aurora, protons, hydrogen, nitrogen

1. Introduction

Understanding the macroscopic behavior of solar and stellar plasmas, and obtaining physical information from analysis of spectra of these plasmas, requires reliable, quantitative information on the fundamental atomic and molecular parameters of the constituent ionic and neutral species. The Ion Trap Laboratory at Appalachian State University has the ability to provide such information, that is, to measure: ion-neutral reaction rate coefficients for processes such as electron capture and proton capture, branching fractions into different excited states for such reactions, collisional de/excitation rate coefficients, unimolecular dissociation rates for ionized molecules, and radiative decay rates of metastable ions.

Currently, we are measuring: (*i*) rate coefficients for proton collisions with H_2 and He, which are important to the coupling of proton and neutral flows in sunspots, and (*ii*) the following parameters of nitrogen important throughout the heliosphere: the radiative lifetime of the ${}^5S_2$ metastable level of N^+, the dissociation rate of N_2^{++}, electron capture rate coefficients by N^+ and N_2^{++} from molecular nitrogen, and the

[*] dawan@appstate.edu, calamaiag@appstate.edu, http://www.phys.appstate.edu
[†] Chemical Physics Program, University of Maryland, College Park, MD 85721, USA
[‡] College of Optical Sciences, University of Arizona, Tucson, AZ 20742, USA

cross section for dissociative electron impact ionization of molecular nitrogen into metastable $^5S_2$ N^+.

Just as atomic physics, plasma physics, and solar physics are intrinsically connected, solar activity and its impact on the heliosphere are intrinsically connected to the study of planetary atmospheres, particularly ionospheres. Radiative decay of metastable $^5S_2$ N^+ results in an emission line doublet at 213.90 and 214.28 nm, an important feature in UV spectra of the Earth's ionosphere which has been referred to as the 'auroral mystery feature', owing to the controversy surrounding its identification and the uncertainties in the atomic parameters (Dalgarno *et al.*, 1981, Bucsela & Sharp, 1989, Meier, 1991, Bucsela *et al.*, 1998). Although the N II] doublet has received considerable attention and some issues have been resolved, there are a number of important issues remaining which will be addressed by this work, as discussed in section 2.1.

Mendoza, Zeippen and Storey (1999) performed a detailed theoretical study of the radiative decay properties of the $^5S_2$ metastable state in the carbon isoelectronic sequence, and concluded, "we would welcome further theoretical and experimental benchmarks that would clarify the inconclusive situation regarding the lifetimes and branching ratios" for the low-charge-state end of this sequence. Our measurement of the radiative decay rate of $^5S_2$ N^+ will thus also help clarify the radiative decay of $^5S_2$ O^{++}, which gives rise to the O III] doublet at 166.08 and 166.61 nm in solar spectra. The intensity of O III] is very sensitive to small variations in electron density, which makes it an important spectroscopic diagnostic of chromosphere-corona transition regions (Del Zanna *et al.*, 2002, Mason & Monsignori Fossi, 1994, Bhatia *et al.*, 1982). The intensity of N II] could also potentially be used as a diagnostic for solar plasmas, but few observations of solar spectra cover the wavelength 214 nm. Vernazza and Reeves (1978), for example, observed the radiative decay of the $^5S_2$ metastable level in carbon-like Ca, Ar, S, Si, Al, Mg, Na, Ne, and F, but O III] and N II] were outside the wavelength range of the observations (28 to 135 nm). Nevertheless, excitation out of $^5S_2$ N^+ gives rise to emission lines such as those at 50.6 and 62.9 nm, which can also be used as a diagnostic. For such diagnostics that involve excitation out of the metastable level, it is the total radiative decay rate, hence the lifetime, for the metastable level's downward transitions that are critical, not the radiative decay rate of the transition observed (*e.g.*, Mason & Monsignori Fossi, 1994). The other critical parameters for such diagnostics are the electron impact de/excitation rate coefficients, which have recently been calculated by Hudson & Bell (2005).

First results for the radiative lifetime of $^5S_2$ N^+ are presented in section 2, as are first results for the dissociation rate of N_2^{++}. Other parameters to be measured are collisional rate coefficients, and the cross section for dissociative electron impact ionization of molecular nitrogen into metastable $^5S_2$ N^+, which is responsible for auroral 214 nm emission. As discussed in section 2.1, estimates for this cross section span a factor of 4. The measurement of this cross section will help resolve whether the primary excitation mechanism for ionospheric N II] is solar photons or auroral electrons (Victor & Dalgarno, 1982, Siskind & Barth, 1987, Cleary & Barth, 1987, Meier, 1991).

To measure atomic and molecular parameters, ions are created in a radiofrequency (RF) ion trap during a 'fill' period by electron impact ionization of gas ad-

mitted into a vacuum chamber. Following the fill period, the number of ions stored is measured as a function of time, and for the nitrogen parameters, the number of UV photons emitted in different bandpasses by radiative decay of metastable states and excited reaction products is also measured as a function of time. A detailed consideration of all collisional and non-collisional rates yields a function, typically an exponential decay, which is fit to the data to determine fundamental atomic and molecular parameters. The nitrogen measurements are discussed in section 2.

The proton-hydrogen and proton-helium measurements, which pertain to the coupling of ion and neutral flows in sunspots and to star formation in the early universe, are discussed in section 3.

The apparatus and method are explained in the context of these measurements, and readers are encouraged to consider parameters required for their research that could be measured using this apparatus. A brief final discussion is given in section 4.

2. Nitrogen Measurements

The nitrogen parameters being measured are:

- The radiative lifetime of the $2s2p^3\ {}^5S_2$ metastable level of N^+
- Rate coefficients for quenching of ${}^5S_2\ N^+$ (and thus N II]) by collisions with N_2
- The cross section for production of ${}^5S_2\ N^+$ by electron impact ionization of N_2
- The unimolecular dissociation rate of the molecular dication N_2^{++}
- Rate coefficients for the electron capture reaction $N_2^{++} + N_2 \rightarrow N_2^+ + N_2^+$
- Branching fractions into different excited states for this reaction

Ions are created in a RF ion trap (see Fig. 1) by electron impact ionization of N_2 gas, and the UV radiation emitted by the stored ion population is then measured as a function of time. The primary source of radiation is the decaying ${}^5S_2$ metastable N^+ ions, which emit photons at a wavelength of 214 nm. In addition to the atomic N^+, doubly-charged molecular ions, or dications, are created and stored, as both ions have the same charge-to-mass ratio. Electron capture by dications from neutral molecules into excited states of singly-charged molecular nitrogen results in emission in bands that overlap with 214 nm (see Fig. 2). The decay rate of these two sources of radiation is measured as a function of nitrogen gas pressure to determine the electron capture rates for both ions, the radiative decay rate of the metastable ${}^5S_2$ N^+, and the dissociation rate of the dications. The photon observations are corrected for ion losses using the technique presented in Daw (2000). The cross section for dissociative electron impact ionization of molecular nitrogen into metastable ${}^5S_2\ N^+$ will be determined relative to the (well-known) cross section for dication production, by comparing the photon rate at 214 nm to the photon rate generated in different bandpasses as a result of dication electron capture, as discussed in section 2.3.

2.1. Motivation

The ${}^5S_2\ N^+$ emission at 214 nm is an important feature in spectra of the Earth's aurora and dayglow that results from dissociative ionization of N_2. Despite the importance of this emission and the attention it has received, interpretation of ionospheric observations is hampered by uncertainties in the atomic and molecular pa-

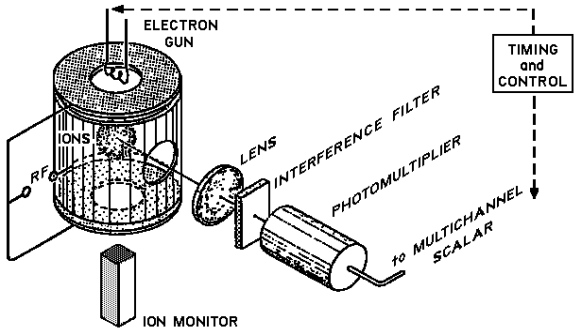


Fig. 1. A diagram showing the ion trap apparatus and elements of the optical detection system.

States of N_2^+ and N^+

Ion	State	r_e(Å)	E (eV above ground)
N_2^+	$C\,^2\Sigma_u^+$	1.26	7.9
	$D\,^2\Pi_g$	1.47	6.4
	$B\,^2\Sigma_u^+$	1.07	3.1
	$A\,^2\Pi_u$	1.17	1.1
	$X^2\Sigma_g^+$	1.12	0
N^+	$2s2p^3\ ^5S_2$		5.80
	$2s^22p^2\ ^1S_0$		4.05
	$2s^22p^2\ ^1D_2$		1.90
	$2s^22p^2\ ^3P_{0,1,2}$		0 - 0.016

205 to 307 nm
138 to 206 nm
214 nm

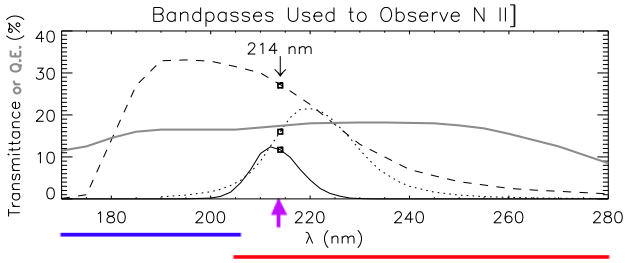


Fig. 2. *Top*: A diagram indicating the observed atomic and molecular transitions on a table of N^+ and N_2^+ states. *Bottom*: A plot of bandpasses used for the observations. The *grey line* is the quantum efficiency of the photomultiplier tube, while the *solid, dotted and dashed lines* are transmittances of different interference filters used for the observations.

rameters. For instance, estimates (mainly deduced from ionospheric observations) of the maximum cross section for electron-impact dissociative ionization into $^5S_2$ N^+ vary between 1×10^{-18} and 4×10^{-18} cm^2 (Dalgarno *et al.*, 1981), and although the energy dependence has been measured (Erdman & Zipf, 1986), knowledge of the absolute value of the cross section has not improved significantly since then. In ref-

erence to ionospheric N II] emission, Bucsela *et al.* (1998) urge caution in adopting atomic and molecular constants that are deduced from observations.

Concerning this problem, significant progress has been made in recent years on the radiative lifetime of $^5S_2$ N^+, both theoretically and experimentally, but significant discrepancies still exist between the experimental results (Träbert *et al.*, 1998) and theoretical results (Brage *et al.*, 1997, Mendoza *et al.*, 1999, Tachiev & Froese Fischer, 2001), and further work is needed to resolve this issue. Thus, the lifetime, excitation rate, and quenching rate measurements for $^5S_2$ N^+, from this work will improve our understanding of ionospheres, of the solar photon and particle fluxes that bombard them, and of the underlying atomic physics.

The N_2^{++} dissociation rate, the rate coefficient for $N_2^{++} + N_2 \rightarrow N_2^+ + N_2^+$, and the branching fraction measurements are perhaps most relevant to the ionosphere of Titan (Lilensten *et al.*, 2005). As the importance of their role in the physics and chemistry of plasmas is realized, the structure and reactivity of molecular dications such as N_2^{++} has been the subject of increasing theoretical and experimental attention in recent years, and methods for understanding these challenging systems are developing (Mathur, 2004, Price, 2003, Cox *et al.*, 2003). For Instance, knowledge of the dissociation rate from the lifetime of N_2^{++} in a storage ring (∼3 s) represents only an order of magnitude estimate (Mathur, 2004). Mathur also indicates that state-resolved studies of low-energy reactions between dications and neutral molecules will help clarify our understanding of dication systems. Moreover, with a significant abundance of N_2^{++} in some ionospheres, the reaction rate for $N_2^{++} + N_2 \rightarrow N_2^+ + N_2^+$ is important to the chemical pathways of such ionospheres.

2.2. Apparatus

Ions are created during a fill period by biasing the electron gun to ∼100 V, and are stored by a combination of RF (∼250 V at 1 MHz) and DC potentials on the ring electrode (see Fig. 1). Of the $m/q = 14$ ions created by electron impact on N_2, 11% are N_2^{++}, and the remainder are N^+ (Halas & Adamczyk 1972). Trap cycles comprise a fill period (typically 5 ms) followed by a data collection period. Light emitted as a result of collisions and by decaying metastable ions is focused by a CaF_2 lens onto an EMR 541Q photomultiplier tube (PMT) operated in photon counting mode. Interference filters are used to select different bandpasses. Photon data is accumulated over an even number of trap cycles. On alternate cycles, the trap is emptied, or 'detuned', by raising the upper end cap voltage and lowering the lower end cap voltage, and counts are subtracted from the photon signal to provide subtraction of the ∼0.5 count/s PMT dark rate and fluorescence of trap components (see Fig. 3). Every few hours, the photon collection is interrupted to collect ion data. To determine the relative number of ions remaining after a given storage time, a dump pulse is applied to the end caps (typically +60 and -60 V), ejecting the ion cloud through the lower end cap and onto a channel electron multiplier (CEM) operated in analog amplification mode. Further details of ion detection may be found in section 3, and data collected with the apparatus may be seen in Fig. 3.

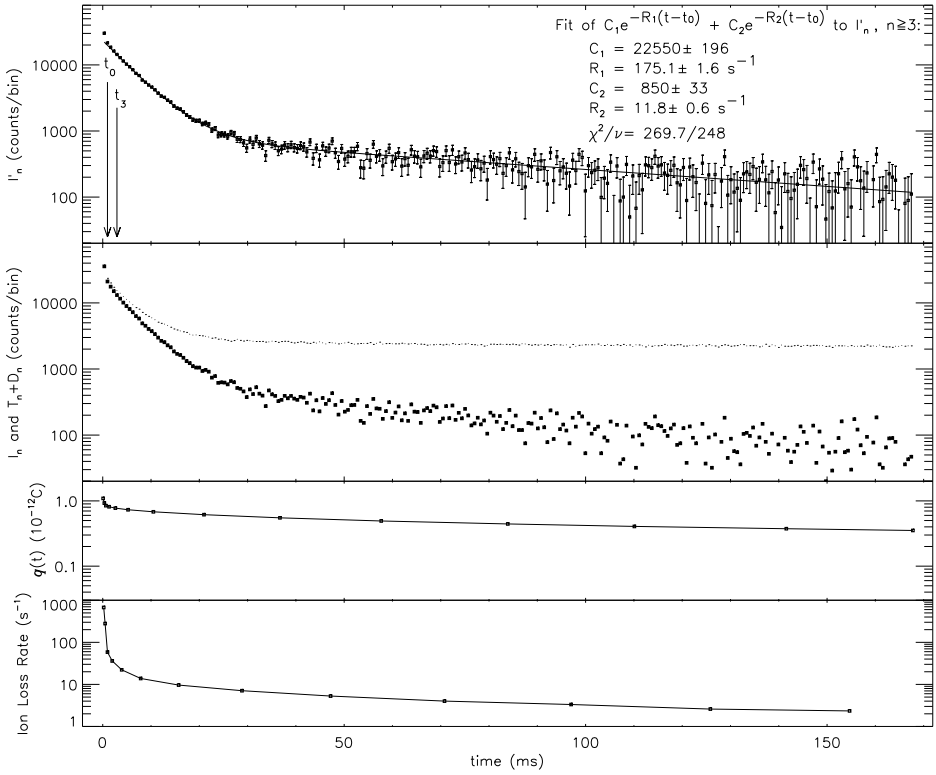


Fig. 3. An example data run comprising 4,781,130 tune/detune cycles of photon collection, taken with N_2 pressure 1.5×10^{-7} torr. *Top*: decay curve obtained by dividing the photon signal by the relative number of ions *vs.* time, as explained briefly here. *Second from top*: For each 0.655 ms bin, the photon signal (*small black squares*) is the number of photons observed during the tuned phase of the cycle minus the number observed during the detune phase: $I_n = T_n - D_n$, where the index indicates bin number. The total $T_n + D_n$ counts per bin are shown by the *tiny dots*, except for the first bin, where $T_n + D_n = 1{,}077{,}357$. Note that n=0 denotes the second bin, as the first bin includes counts from excited neutrals remaining briefly in the field of view after the end of the fill period. *Third from top*: For every photon collection interval (typically 10^5 T/D cycles), the amplified ion cloud charge, $q(t)$, was measured for a number of values of t. The *small black squares* show the average of the data taken for all photon collection intervals. The relative number of ions for each bin is determined by interpolating between the $q(t)$ measurements and is given by $Q(t_n) = q(t_n)/q(t_0)$, so that $I'_0 = I_0$. *Bottom*: Ion loss rate shown for discussion. *Back to the Top*: so explicitly, each point in the decay curve is given by $I'_n = I_n/Q(t_n)$ and has an uncertainty $\sigma_n = \sqrt{T_n + D_n}/Q(t_n)$. Equation 6, specifically, $I'(t - t_0)$, was fit to the data skipping the first 2.62 ms to allow the ion detection system to recover from the fill period, with the results shown above.

2.3. Data and Analysis

The method used for these measurements includes a number of extensions of the method developed to measure the radiative lifetime of the $^1S_0$ metastable level of Ne^{++} (Daw, 2000). In this method, a decay curve is defined as $I'(t) = I(t)/Q(t)$, where $Q(t) = N(t)/N(0)$ and $N(t)$ is the number of ions stored in the trap. $I(t)$ denotes the observed photon signal rate. That is, a decay curve is the photon signal divided by the relative number of ions, as shown in Fig. 3. A detailed consideration

of all populations and processes in the trap yields a function that can be fit to the decay curves to determine fundamental parameters for the rates of the processes, as discussed below. Basically, there are two decay components in the data: a fast decay component, due to decaying metastable ions, and a slow decay component, due to excited N_2^+ produced as a result of N_2^{++}+N_2 collisions. Thus, the data is well-described by the sum of two exponential decays.

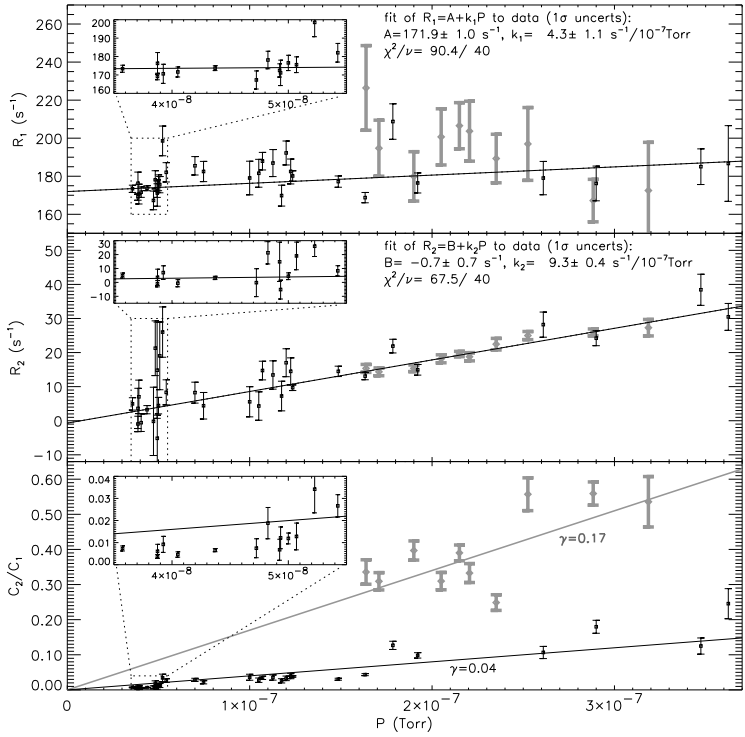


Fig. 4. Decay curves were collected for N_2 pressures ranging between 3×10^{-8} and 4×10^{-7} torr. The simplified form of $I'(t)$ (equation 6) was fit to the decay curves, yielding values for R_1 and R_2 *vs.* pressure, and values for the ratio C_2/C_1, which is a function of the photon detection bandpass, fill period, and pressure. The data above was collected using two different interference filters: one with a 11 nm FWHM bandpass centered at 213 nm (*black data*), and one with a 20 nm FWHM bandpass centered at 220 nm (*thick grey data*). Both transmittances are shown in Fig. 2. Linear fits to R_1 and R_2 as a function of P yielded results for: the radiative lifetime τ of $^5S_2$ N^+, the rate coefficient k_1, the dissociation rate B of N_2^{++}, and the rate coefficient k_2'.

The photon signal includes two sources of photons: radiative decay of metastable $^5S_2$ N^+, and collisions of $N_2{}^{++}$ with N_2.

$$I(t) = I_1(t) + I_2(t) \tag{1}$$

The signal rate for radiative decay photons is given by

$$I_1(t) = \epsilon_1 A N_1(t), \tag{2}$$

where ϵ_1 is the detection efficiency for the radiative decay photons, A is the total decay rate (or sum of A-values) for the radiative transitions being observed, and $N_1(t)$ is the number of metastable ions.

The signal rate for the collisionally-produced photons is given by

$$I_2(t) = \epsilon_2 k_2 P N_2(t), \tag{3}$$

where ϵ_2 is the probability (per collision) of detecting a photon, k_2 is the rate coefficient for the reaction $N_2 + N_2^{++} \rightarrow N_2^+ + N_2^+$, P is the N_2 pressure (converted to particle density for room temperature gas if k has units of cm^3/s), and $N_2(t)$ is the number of N_2^{++} ions. In detail, ϵ_2 is the probability distribution (in wavelength) for the photons generated by the collisions convolved with the bandpass of the photon detection system. This reaction has an exoergicity of 11.5 eV for ground-state reactants and products, and can easily produce two photons at wavelengths near 214 nm if both product ions are excited.

Because N^+ and $N_2{}^{++}$ have the same charge-to-mass ratio, the relative number of ions measured as a function of time reflects the combined population (both species). The functional form for the decay curve is obtained for this situation from a detailed consideration of rates for all collisional and non-collisional processes by solving the differential equations describing the level populations of both species.

Defining the following symbols:

$\tau =$ the radiative lifetime of $^5S_2$ N^+ (the reciprocal of the sum of the A-values for the 214 nm doublet)

$B =$ the dissociation rate of N_2^{++}

$k_0 =$ the rate coefficient for loss of N^+ ions via collisions with N_2

$k_1 =$ the difference between the ion-neutral collisional loss rate coefficient for the metastable level and the weighted average for the total ion population, plus the sum of rate coefficients for collisional de-excitation of the metastable level

$k_2 =$ the rate coefficient for the reaction $N_2^{++}+N_2 \rightarrow N_2^+ + N_2^+$

$k_3 =$ the ion-neutral collisional loss rate coefficient for N_2^{++} by all channels other than the reaction $N_2^{++}+N_2 \rightarrow N_2^+ + N_2^+$

$k_4 =$ the rate coefficient for creation of N^+ from N_2^{++} + N_2 collisions

$P =$ N_2 pressure (for k's with units of s^{-1}/torr), or particle density (for cm^3/s),

the fraction of N_2^{++} ions in the total ion population is given by

$$f(t) = [(f_0^{-1} - \delta)e^{R_2 t} + \delta]^{-1}, \tag{4}$$

where $R_2 = B + k_2' P$, $k_2' = k_2 + k_3 - k_0$, $\delta = (k_2' P - k_4 P - B)/R_2$, and $f_0 \equiv f(0)$. Note that the processes described by k_3 and k_4 are expected to not be significant, but in the final analysis, the impact of all processes on the measured parameters will be evaluated (including some processes omitted here for the sake of clarity). A full derivation is beyond the scope of this article, but a similar derivation may be found in Daw (2000). Here, the functional form for the decay curve is given by

$$I'(t) = C_1 e^{-R_1 t + R_2 t} f(t)/f_0 + C_2 f(t)/f_0 \tag{5}$$

where C_1 and C_2 are the detected photon rates at t=0 for the metastable decay and collisionally produced photons, and $R_1 = \tau^{-1} + k_1 P$.

For f_0=0.11, equation 5 for $I'(t)$ is very close in form to the sum of two exponentials:

$$I'(t) = C_1 e^{-R_1 t} + C_2 e^{-R_2 t} \tag{6}$$

with the decay rate R_1 pertaining to the metastable decay component of the decay curve, and the decay rate R_2 pertaining to the component produced by N_2^{++}+N_2 collisions.

Decay curves were collected for a number of N_2 pressures, and for the preliminary results, equation 6 was fit to the data to determine τ, k_1, B, and k_2' (see Fig. 4). For the final results, the complete form of $I'(t)$ will be fit to all decay curves simultaneously to determine additional parameters such as k_3 and k_4. Note that the rate coefficient k_0 for loss of N^+ via collisions with N_2 can be estimated from the measured loss rate of the (primarily N^+) ion signal, and since k_2' is much greater than k_3 and k_0, the rate coefficient for the reaction $N_2 + N_2^{++} \rightarrow N_2^+ + N_2^+$ can be determined by $k_2 = k_2' - k_3 + k_0$. Branching fractions into different excited states of N_2^+ will be determined by observing the corresponding emission bands of N_2^+. Finally, the production cross section for $^5S_2$ N^+ can be determined because the ratio of $^5S_2$ N^+ to N_2^{++} ions at t=0 is given by $(C_1/C_2)(\epsilon_2/\epsilon_1)(\tau k_2 P)$.

With this preliminary data, the result for the radiative lifetime of $^5S_2$ N^+ is 5.8 ± 0.1 ms, and the result for the dissociation rate of N_2^{++} is less than 1 s^{-1}. The rate coefficients are not presented yet, as the ion gauge was not calibrated for this preliminary work. Nevertheless, the data demonstrates the viability of the measurement method. For future data collection, a calibrated ion gauge will be used, amongst other improvements to the apparatus, resulting in lower uncertainties. For instance, we expect to measure the $^5S_2$ N^+ lifetime with an uncertainty of less than 1%, to help clarify the still inconclusive situation regarding this well-studied lifetime. The other atomic and molecular parameters to be measured will provide valuable information for ionospheric processes that currently have order-of-magnitude uncertainty associated with them. The cross section for production of $^5S_2$ N^+ by electron impact on N_2 currently has a factor of four uncertainty, and our measurement will help resolve whether the primary excitation is by solar photons or by electrons.

3. Proton Measurements

We are currently measuring rate coefficients for proton collisions with H_2 and He for their importance to the coupling of proton and neutral flows in sunspots, and for their importance to star formation in the early universe. Collision energies between protons and molecules in the experiment are typically a few eV or less and are controlled by the RF potential. At these energies, H^+ + He collisions are elastic, as no excited states are accessible. For H^+ + H_2 collisions, the charge exchange reaction $H_2 + H^+ \rightarrow H_2^+ + H$ has a threshold of 1.83 eV and both inelastic and elastic collisions occur. Since molecular hydrogen is observed in sunspots (Schüehle *et al.*, 1999) and starspots (Wood & Karovska, 2004), these collisional rates are important to the coupling of plasma and neutral flows.

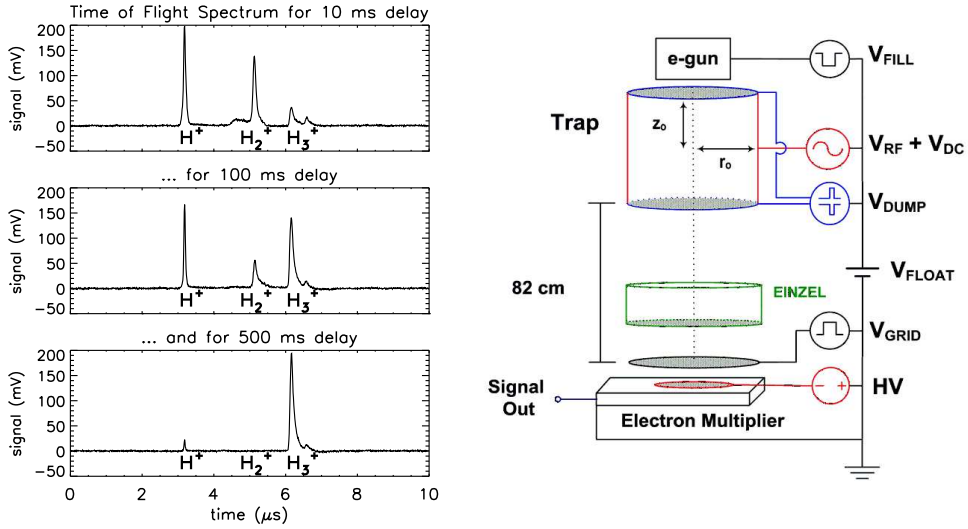


Fig. 5. *Left*: Time of flight spectra obtained when the trap was tuned to store H_2^+, shown for three different delay times after ions are created by electron impact. Note that the ion extraction was optimized for protons, so that protons arrive in a single peak, while the structure seen in the other ions' time of flight distributions provides valuable information on the ion temperatures in the trap. *Right:* A diagram of the apparatus and applied voltages. The trap dimensions are r_0=1.76 cm, z_0=1.86 cm.

The charge exchange process $H_2 + H^+ \rightarrow H_2^+ + H$ is predicted to be the dominant destruction mechanism of H_2 during the epoch of first star formation. However, the collision rate coefficient differs among published calculations by orders of magnitude (Savin *et al.*, 2004). Collision rate coefficients can be measured by observing the loss of the reactant(s) and or the growth of the product(s). In this case, the product H_2^+ reacts rapidly with H_2 to form H_3^+, as seen in Fig. 5. Also shown in this figure is the apparatus for the proton-hydrogen and proton-helium collision rate measurements, which is similar to that used for the nitrogen measurements, except that ion detection alone is sufficient for these measurements, and a longer flight path is used to obtain better charge-to-mass resolution. The reaction $H_2^+ + H_2 \rightarrow H_3^+ + H$ is being studied for its importance to ionospheric chemistry, and for its effect on the H_2^+ ions produced by $H_2 + H^+$ collisions in the ion trap. In the future, similar such reactions between ions and molecules can be studied.

4. Results and Discussion

The results from the preliminary nitrogen data collection are: 5.8 ± 0.1 ms for the radiative lifetime of $^5S_2$ N^+, and less than 1 s^{-1} for the dissociation rate of N_2^{++}. Further data collection will provide lower uncertainties for these parameters, and measurements of additional parameters such as the cross section for production of $^5S_2$ N^+ by electron impact on N_2, which will help to resolve whether the primary ionospheric excitation of N II] is by solar photons or electrons, and collision rate

coefficients for H^++H_2 and H^++He collisions important to coupling of plasma and neutral flows in sunspots. A number of additional measurements are in progress, and researchers are encouraged to request and/or collaborate on the measurement of parameters relevant to their work.

Acknowledgements. We would like to thank all members of the local and scientific organizing committees for providing this invaluable opportunity to share scientific ideas, initiate collaborations, and conduct observations of the eclipse. We would like to thank Harvard University and the Smithsonian Astrophysical Observatory for resources toward initial exploration of the nitrogen work, and in particular, Alex Dalgarno, William H. Parkinson, and Peter L. Smith for informative discussions. Dr. Daw would like to thank the University of Hawaii IfA eclipse group, the many at Appalachian State University who supported this opportunity to collaborate on eclipse observations with the IfA group, and the people of Libya for their welcoming hospitality, support, and blessings of these scientific enterprises. The work on hydrogen is supported by NSF grant AST-04-06706 to Appalachian State University, NASA grant NAG5-5420 to Columbia University, and NSF grant AST-03-07203 to Columbia University. The work on nitrogen is supported by Research Corporation award CC6409 to Appalachian State University.

References

Bhatia, A. K., Kastner, S. O., and Behring, W. E.: 1982, *ApJ*, **257**, 887
Brage, T., Hibbert, A., and Leckrone, D.S.: 1997, *ApJ*, **478**, 423
Bucsela, E. J., Cleary, D. D., Dymond, K. F., and McCoy, R. P.: 1998, *J. Geophys. Res-Space Phys*, **103**, 29215
Bucsela, E. J. and Sharp, W. E.: 1989, *J. Geophys. Res*, **94**, 12069
Cleary, D. D. and Barth, C. A.: 1987, *J. Geophys. Res*, **92**, 13635
Cox S. G., Critchley A. D. J. , Kreynin P. S. , McNab I. R. , Shiell R. C., and Smith F. E.: 2003, *Phys. Chem. Chem. Phys.*, **5**, 663
Dalgarno, A., Victor, G. A., and Hartquist, T. W.: 1981, *Geophys. Res. Lett.*, **8**, 603
Daw A.: 2000, *Ph. D. thesis*, Harvard University
Daw A., Parkinson W. H., Smith P. L., and Calamai A. G.: 2000, *ApJ*, **533**, L179‘
Del Zanna, G., Landini, M., and Mason, H. E.: 2002, *A&A*, **385**, 968
Erdman, P.W., and Zipf, E.C.: 1986, *J. Geophys. Res-Space Phys.*, **91**, 1345
Halas, St., and Adamczyk, B.: 1972, *Int. J. Mass Spectrom. Ion Phys.*, **10**, 157
Hudson, C. E. and Bell. K. L.: 2005, *A&A*, **430**, 725
Lilensten J., Witasse O., Simon C., Solidi-Lose H., Dutuit O., Thissen R., and Alcaraz C.: 2005, *Geophys. Res. Lett.*, **32**, L03203
Mason, H. E., and Monsignori Fossi, B. C.: 1994, *A&A Rev.*, **6**, 123
Mathur, D.: 2004, *Phys. Rep.*, **391**, 1
Meier, R. R.: 1991, *Space Sci.Rev.*, **58**, 1
Mendoza, C., Zeippen, C. J., and Storey, P. J.: 1999, *A&AS*, **135**, 9
Price, S. P.: 2003, *Phys. Chem. Chem. Phys.*, **5**, 1717
Savin, D. W., Krstic, P. S., Haiman, Z. N., and Stancil, P. C.: 2004, *ApJ*, **607**, L147
Schüehle, U., Brown, C. M., Curdt, W., and Feldman, U.: 1999, in J. C. Vial, and B. Kaldeich-Schümann (Eds.), ESA SP-446: Proceedings of the 8th SOHO Workshop
Siskind, D. E. and Barth, C. A.: 1987, *Geophys. Res. Lett.*, **14**, 479
Tachiev, G., and Froese Fischer, C.: 2001, *Can. J. Phys.*, **79**, 955
Träbert, E., Wolf, A., Pinnington, E. H., Linkemann, J., Knystautas, E. J., Curtis, A., Bhattacharya, N., and Berry, H. G.: 1998, *Phys.Rev.A*, **58**, 4449
Vernazza, J. E. and Reeves, E. M.: 1978, *ApJS*, **37**, 485
Victor, G. A. and Dalgarno, A.: 1982, *Geophys. Res. Lett.*, **9**, 866
Wood, B. E. and Karovska, M.: 2004, *ApJ*, **60**, 502


ving time available:
tion to particular solar events
or for projects requiring long
development of new
is small and almost constant over
easy to correct – depends on

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

SOLAR RESEARCH PROGRAMS AT IRSOL, SWITZERLAND

RENZO RAMELLI and MICHELE BIANDA*
Istituto Ricerche Solari Locarno, Via Patocchi 57, CH-6005, Locarno, Switzerland

JAN O. STENFLO†
Institute of Astronomy, ETH Zentrum, CH-8092 Zurich, Switzerland

and

PHILIPPE JETZER
Institute for Theoretical Physics, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Abstract. The Zurich IMaging POLarimeter (ZIMPOL) developed at ETH Zurich and installed permanently at the Gregory Coudé Telescope at Istituto Ricerche Solari Locarno (IRSOL) allows a polarimetric precision down to 10^{-5} to be reached. This makes it possible to perform several accurate spectro-polarimetric measurements of scattering polarization and to investigate solar magnetic fields through the signatures of the Hanle and Zeeman effects. The research programs are currently being extended to monochromatic imaging of the Stokes vector with a recently installed Fabry-Perot rapidly tunable filter system with a narrow pass band of about 30 mÅ. The spatial resolution is being improved by the installation of an Adaptive Optics system.

Key words: solar physics, polarimetry, magnetic fields

1. Introduction

The great advances in high precision polarimetry that have been achieved with the introduction of the Zurich IMaging POLarimeter (ZIMPOL) a decade ago opened a new window in solar physics. Polarimetry is in fact a very powerful tool that can be used to study solar magnetic fields as well as the physical processes behind the generation of polarization in atomic and molecular spectral lines. Magnetic field measurements through Zeeman effect signatures, which appear in the presence of strong and oriented magnetic fields, have long been performed at many observatories. With the high polarimetric precision of ZIMPOL it has become possible to extend the magnetic field diagnostics to weak fields and to fields which are tangled on scales below the spatial resolution, which are invisible to the Zeeman effect but get revealed by the Hanle effect (Hanle, 1924) (for details see Trujillo Bueno, 2006).

Spectro-polarimetry is currently the main field of research at the Istituto Ricerche Solari Locarno (IRSOL). Advantage is taken from the circumstance that a ZIMPOL system is permanently installed at IRSOL. In addition the Gregory Coudé Telescope (GCT) of the observatory is very well suited for polarimetric measurements, since the amount of instrumental polarization is low and stays practically constant during the observing day, since it is a function of declination only. Therefore it can easily

* Also affiliated with ETH Zurich, CH-8092 Zurich, Switzerland
† also affiliated with the Faculty of Mathematics & Science, University of Zurich

Fig. 1. IRSOL Observatory. In the left part of the building there is the observing room, while the spectrograph room is underground. All is painted white with titanium dioxide. The institute is located 500 m above sea-level, above the city of Locarno, Switzerland.

be accounted for. With a Fabry-Perot filter system and an adaptive optics system recently installed at IRSOL we plan to start several new interesting projects.

2. IRSOL - The Institute

The observatory at the Istituto Ricerche Solari Locarno (IRSOL) (Figure 1), located in southern Switzerland, was constructed in 1960 by the Universitäts-Sternwarte Göttingen (USG), Germany. In 1984, after USG moved its observing activity to the facilities at Observatorio del Teide on Tenerife, a local foundation (FIRSOL) acquired the observatory in Locarno. The partially dismantled instrumentation was rebuilt and improved, in collaboration with USG, University of Applied Sciences of Wiesbaden (Germany), and the Institute of Astronomy at ETH Zurich. The scientific collaboration with ETH Zurich allowed the implementation at IRSOL of an important polarimetry observing program, first with a beam exchange polarimeter and then with ZIMPOL.

3. Instrumentation at IRSOL

The IRSOL telescope (Figure 2 and 3) is a 45 cm aperture Gregory Coudé with 24 m effective focal length. The field stop at the prime focus reduces the field of view to a 200 arcsec diameter circular image. The rest of the solar image is reflected away from the main light beam. This reduces heating and scattered light and is of particular advantage when observing low intensity structures like sunspots, spicules and prominences. The relative orientation of the two folding mirrors M3

and M4 (Coudé) changes only with declination and is orthogonal at the time of the equinoxes. As a consequence the instrumental polarization, originating through oblique reflections, is almost constant during the day and virtually vanishes during the equinoxes (Sánchez *et al.*, 1991). A Gregory Coudé telescope is thus very well suited for polarimetric measurements.

Fig. 2. The 45 cm aperture Gregory Coudé vacuum telescope.

An automatic guiding system developed by the University of Applied Sciences Wiesbaden (Küveler *et al.*, 1998) is also available. Its operation is based on the solar image obtained from the light rejected by the field stop at the primary focus.

The Czerny-Turner spectrograph with 10 m focal length is based on a 180 × 360 mm grating with 300 lines per mm and 63° Blaze angle. A prism based predisperser allows to select the spectral band entering in the spectrograph without overlap of the grating orders.

Monochromatic imaging observations of the solar surface can be performed using the recently installed Fabry-Perot filter system in collimated configuration (Feller *et al.*, 2006) (Figure 4). The system uses two temperature controlled lithium niobate etalons with an aperture of 70 mm. The transmitted wavelength can be selected by electrically tuning the refractive index of the etalon medium, by varying the temperature, or by tilting the etalon. The bandwidth is about 30 mÅ.

An adaptive optics (AO) system based on a tip-tilt mirror and a 37 actuator deformable mirror is currently being installed and tested in collaboration with the University of Applied Sciences of Southern Switzerland, SUPSI, and with ETH Zurich. The system follows the design of the infrared AO system installed at the McMath-Pierce Solar Telescope at Kitt Peak (Keller *et al.*, 2003). The first tests made with the tip-tilt mirror have already given good results (see Figure 5).

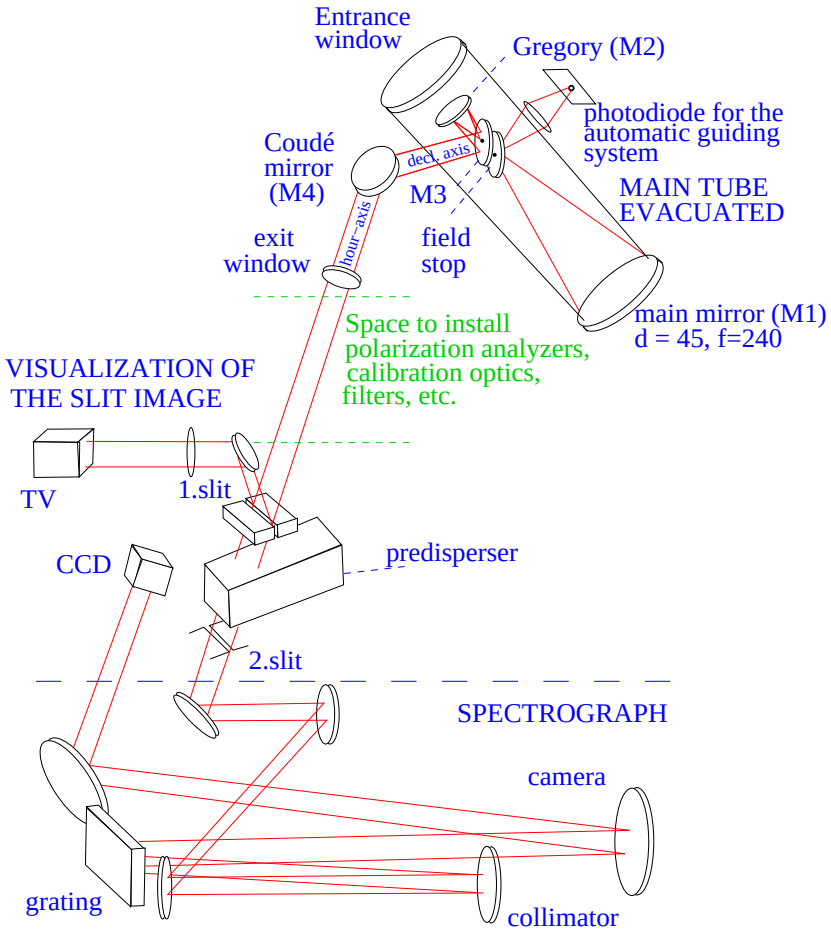


Fig. 3. Optical scheme of the Gregory Coudé Telescope with the spectrograph.

Two polarimeters are available. The oldest one operating at IRSOL is the dual beam exchange device based on a Savart plate and rotating quarter and half wave retarder plates (Bianda *et al.*, 1998). A polarimetric precision of a few 10^{-4} can be reached, but it can be affected by seeing-induced cross-talk, because the technique requires two exposures taken at different times. The second polarimeter is ZIMPOL (Povel, 1995; Gandorfer *et al.*, 2004), which is installed permanently at IRSOL since 1998. Its main advantage is that it is free from seeing-induced effects thanks to its high modulation rate: 42 kHz (obtained with a piezoelastic modulator), or 1 kHz (obtained with ferro-electric liquid crystal modulators). Another advantage is that the same pixels of the CCD ZIMPOL sensor are used to measure all Stokes parameters. Therefore the Stokes Q/I, U/I and V/I images are not influenced by different pixel efficiencies. The ZIMPOL polarimetric accuracy depends mainly on

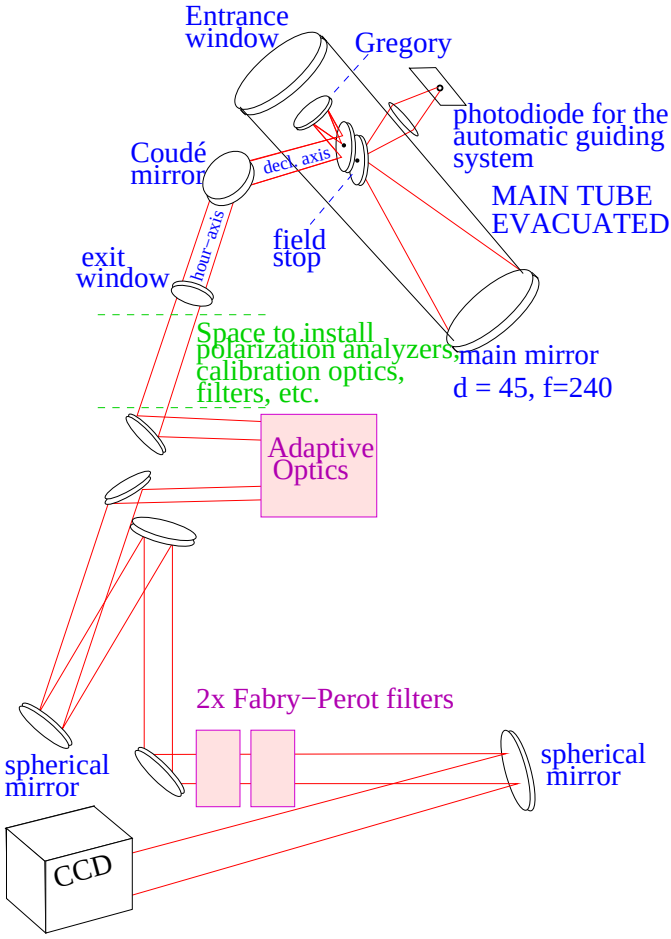


Fig. 4. Optical scheme of the Gregory Coudé Telescope for measurements with the Fabry-Perot filter system.

the photon statistics. With long exposure times it has already been possible at IRSOL to reach an accuracy of about 10^{-5}.

4. Scientific research projects at IRSOL

The research projects at IRSOL take advantage of the very good polarimetric and spectral accuracy of the instrumentation. A large amount of observing time is available to carry out monitoring measurements or for projects requiring long observing times (which cannot easily be done at large telescope facilities, where the observing time is shared by different research groups according to a predefined program). In addition it is possible to be very flexible with the programs to allow fast reaction to

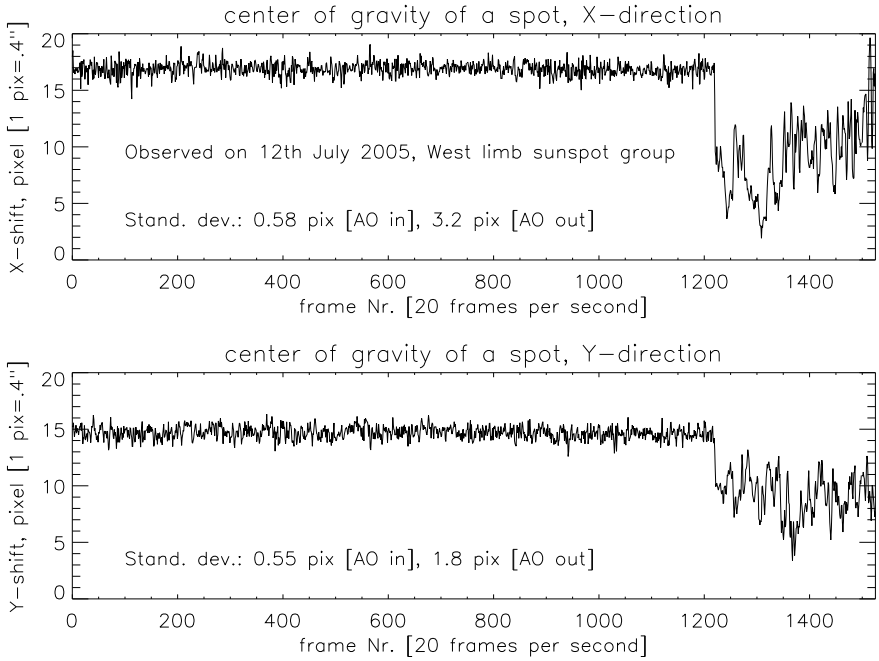


Fig. 5. First tests with the tip-tilt correction of the AO system. The two plots show the displacement of the barycentre of a sunspot in the x resp. y direction with and without tip-tilt correction. The tip-tilt mirror was switched off around frame number 1200.

particular solar events. The modular layout of the instrumentation inside the observing room is very convenient for installation and testing of new instrumentation.

Examples of the scientific results obtained at IRSOL in recent years are:

- Investigations of the Hanle effect in the quiet chromosphere (Bianda *et al.*, 1999).
- Publication of the first two volumes of the "Atlas of the Second Solar Spectrum" (Gandorfer, 2000; Gandorfer, 2002).
- Discovery of vast amounts of hidden magnetism in the solar photosphere (Trujillo Bueno *et al.*, 2004; Stenflo, 2004).
- Determination of novel constraints on impact polarization in solar flares (Bianda *et al.*, 2005).
- Measurements of full Stokes profiles in prominences in H_α (Figure 6), He D_3, H_β, and in spicules in He D_3 (Ramelli and Bianda, 2005; Ramelli *et al.*, 2005).
- First polarimetric measurements of the molecular Zeeman effect in several CH lines in the G band (Asensio Ramos *et al.*, 2004) and of the Paschen-Back effect in CaH transitions (Berdyugina *et al.*, 2006).

Different other observing programs are also foreseen in the future. They will focus on solar magnetism and polarimetry with the Fabry-Perot filter system or with the spectrograph. They also include synoptic type programs (eg. variations of the Hanle-effect signatures with heliographic latitude and solar cycle). Furthermore IRSOL is open to coordinated type programs with other observatories: simultaneous

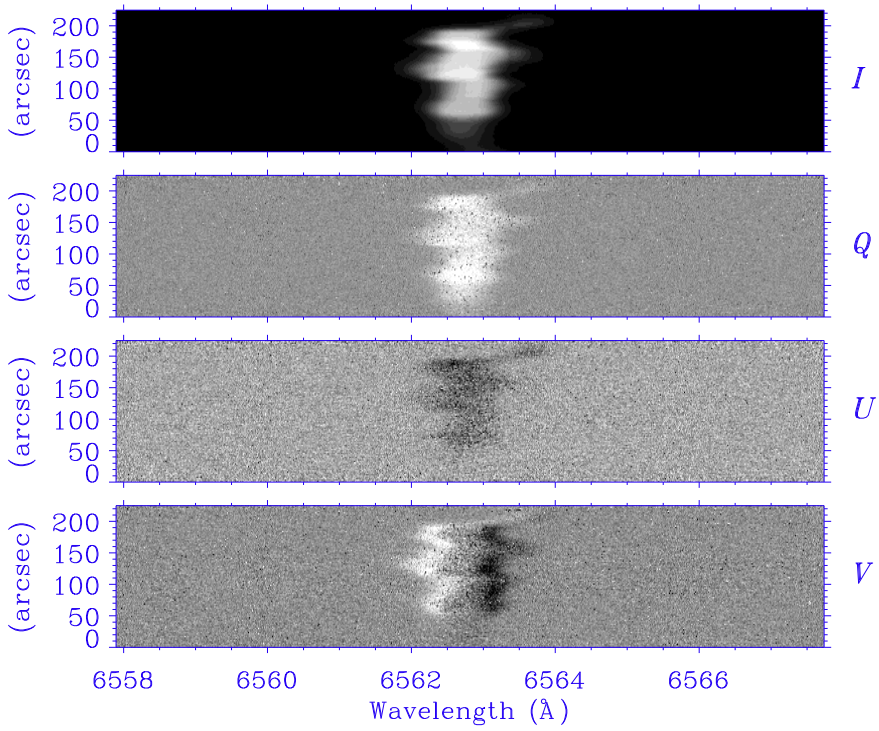


Fig. 6. Example of full Stokes spectro-polarimetric recording of an erupting prominence in H_α. The measurement was taken on 27th August 2004 at about 50 arcseconds form the limb.

observations of solar features with complementary sets of instruments or supporting type observations that complement the science of another project.

Acknowledgements. We are grateful for the financial support provided by the canton of Ticino, the city of Locarno, ETH Zurich and the Fondazione Aldo e Cele Daccò. We also appreciate the support by the organizers of the SPSE conference.

References

Asensio Ramos A., Trujillo Bueno J., Bianda M., Manso Sainz R., and Uitenbroek H.: 2004, *ApJ*, **611**, L61
Berdyugina, S.V., Fluri, D.M., Ramelli, R., Bianda, M., Gisler, D., and Stenflo, J.O.: 2006, *ApJ*, **649**, L49
Bianda, M., Solanki, S.K., and Stenflo, J. O.: 1998, *A&A*, **331**, 760
Bianda, M., Stenflo, J.O., and Solanki, S. K.: 1999, *A&A*, **350**, 1060
Bianda, M., Benz, A.O., Stenflo, J.O., Kueveler, G., Ramelli, R.: 2005, *A&A*, **434**, 1183
Feller, A., *et al.*: 2006, Proceedings of the Fourth Solar Polarization Workshop, Boulder 2005.
Gandorfer, A.: 2000, VdF, ISBN 3-7281-2764-7
Gandorfer, A.: 2002, VdF, ISBN 3-7281-2844-4
Gandorfer, A., *et al.*: 2004, *A&A*, **422**, 703
Hanle, W.: 1924, *Z. Phys.*, **30**, 93
Keller, C. U., Plymate, C. and Ammons, S. M.: 2003, Proceedings of the *SPIE*, **4853**, 35
Küveler et al.: 1998, *Sol. Phys.*, **182**, 247

Povel, H.P.: 1995, *Opt. Eng.* **34**, 1870
Ramelli, R. and Bianda, M.: 2005, *Astronomy and Astrophysics Space Science Library*, **320**, 215
Ramelli, R., Bianda, M., Trujillo Bueno, J., Merenda, L., and Stenflo, J. O.: 2005, in D. E. Innes, A. Lagg, and S. A. Solanki (Eds.), ESA SP-596: Chromospheric and Coronal Magnetic Fields
Sánchez Almeida, J., Martínez Pillet, V., and Wittmann, A. D.: 1991, *Solar Phys.* **134**, 1
Stenflo, J. O.: 2004, *Nature*, **430**, 304
Trujillo Bueno, J., Shchukina, N., and Asensio Ramos, A.: 2004, *Nature*, **430**, 326
Trujillo Bueno, J., these proceedings

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

THE ISLAMIC INFLUENCE ON WESTERN ASTRONOMY

H. NUSSBAUMER
Institute of Astronomy, ETH Zurich, Switzerland.

Abstract. Within one hundred years of the Prophet's death in 632, Islam had conquered the lands from the Indus to the Atlantic. The new culture was eager to assimilate what was best in the heritage of the conquered. Science, and in particular astronomy and astrology, were given very high priority. Islamic scholars soon realized that Ptolemy represented the scientifically most advanced aspects of ancient astronomy. Islamic scientists critically developed Ptolemy's theories, they constructed new astronomical instrumentation, improved observations, and created algebra and spherical trigonometry for the interpretation of the observations. From India they introduced the positional decimal system including zero, which from Northern Africa and Spain found its way to Europe. Astronomical knowledge in the Christian West was at the same time stuck at a very primitive level. This began to improve when towards the end of the 10th century contacts between the Islamic and Christian cultures, particularly in Spain, began to multiply. In the 11th and 12th century intellectual centres, Toledo housing the most important one, translated all the available astronomical works from Arabic into Castilian or Latin. Up to the days of Copernicus these works formed the basis on which European astronomy was building and developing.

Key words: History of astronomy, Islamic astronomy, Middle Ages

1. Introduction

The Prophet died in the year 632 A.D. Immediately after his death the Islamic movement succeeded in becoming a world power, a universal religion, and a new civilization, stretching from India to the Atlantic Ocean.

The Islamic conquest was different from what we usually associate with conquest and domination. There was respect for the cultural heritage of the conquered and a keen wish to assimilate what they deemed worthy. This was particularly true in respect to the old Greek civilization. The Islamic civilization emerged from a melting pot of Greek, Indian, Persian, Mesopotamian, and North African heritages growing on a fertile Arabic ground. This new Islamic civilization had first an indirect and from the 11th to the 14th century a very direct influence on the history of Europe.

2. The remains of former glory

What was the scientific environment of the lands conquered by Islam? Let us have a brief glance at history.

In 338 Phillip of Macedonia conquered Greece. His son, Alexander the Great, extended the Empire to reach from Egypt to the Caspian sea and from the Mediterranean to the Indus. Greek culture was spread to all these places, after all, Alexander had been a pupil of Aristotle. He founded Alexandria in 331 B.C. as a new centre of the Empire. With the establishment of the Museion around 295 B.C. the successors of Alexander intended to create an Academy along the one of Aristotle. Euclid,

Apollonius, and Erathostenes worked in Alexandria, and in 150 A.D. Ptolemy published the Almagest.

In the meantime Egypt and the whole Mediterranean had become a Roman dominion and Christianity was spreading. With the advent of Christianity science became increasingly suspect. The assassination of Hypatia is a testimonial of the Christian fight against science. Hypatia was a learned woman of high reputation who is reported to have taught mathematics and astronomy. She was brutally killed in 415 by a Christian fundamentalists mob. For those who killed her, knowledge was not to be gained through scientific research, but was granted through divine revelation, and the Church had the monopoly on that divine transfer.

However, in some places like Alexandria in Egypt, Antioch in Syria, Edessa in Mesopotamia, or Byzantium the Hellenistic culture was still present, either in academies or half forgotten in libraries.

3. The beginning of Islamic science

Shortly after 640 the Sasanide empire was conquered by Islam. Baghdad was founded in 762 as the new capital by the second Abbasid caliph al-Mansur. And here in Baghdad we see the first important impetus of Islamic science. It began with al-Mansur, who was much interested in astrology, and it continued with his successors, e.g. Harun al-Rashid, who was Caliph from 786 to 809. It must have been a truly refined society. The splendour of the stories of Thousand and One Night bear witness. Caliph al-Mamun, the son of Harun al-Rashid, was particularly keen on collecting ancient sources. Much has been said about the Bayt al-hikma, the House of Wisdom. Its roots lie probably in the archive of al-Mansur, where many Persian documents were translated into Arabic. Particularly at the times of al-Mansur it may also have functioned as a gathering place for intellectuals, and a kind of academy (Gutas, 1998; Kennedy, 2005). In that place and in the households of the upper social classes Christian, Jewish, Muslim and Pagan intellectuals worked and translated from Greek, Syriac, and Persian into Arabic all that had survived of the philosophical and scientific traditions. The biggest treasure for astronomy was Ptolemy's Almagest. An excellent mathematical background was provided by the works of Euclid, Archimedes, and Apollonius. These translations became part of the Islamic culture. Arabic became an international language of science, and the dissemination of knowledge was further facilitated by the use of paper which around 750 was introduced from China.

4. Islamic contributions to astronomy

4.1. Observations

The classical astronomical observations of Aristarch, Hipparchos, and Ptolemy were critically repeated by Islamic observers. To mention just a few: Al-Battani (850-929) looked at the length of the year, he found changes in the inclination of the ecliptic and the rate of precession. In contrast to Ptolemy, who thought that the solar apogee was fixed, al-Biruni (973-1048) found that it slowly moves. Al-Sufi (903-986) published in 964 the 'Book of Fixed Stars'. It was an update of Ptolemy's

list, with a more accurate determination of the brightness of the stars. It was widely used, and served as source of information when the Alfonsine Tables were set up in the 13th century. - It is strange that the Islamic astronomers did not discover the supernova of 1054, whereas plenty of observations are recorded from China.

Islamic scientists were the first to create astronomical observatories as scientific institutions. Maraghah in Persia was established by al-Tusi in 1262. That observatory was much more than a mere centre of observation. It possessed a fine library with books on a wide range of scientific topics. Theoretical work in mathematics, astronomy, and philosophy was vigorously pursued. The Samarkand observatory of Ulugh Beg was in operation from 1420 to 1449. Ulugh Beg's Catalogue of stars, is the first comprehensive stellar catalogue since that of Ptolemy.

Istanbul observatory functioned from 1575 to 1580, thus practically at the same time as the one of Tycho Brahe. However, there are no indications that they knew of each other. It was built 1575-1577 for Taqi al-Din by the Sultan Murad III. The reason for its closure is uncertain. Did religious leaders persuade al-Din that prying into the secrets of nature would only bring misfortune, or did he fail with his astrological predictions? Tycho Brahe lost his observatory because the new Danish king was not interested in astronomy.

4.2. Instrumentation

In addition to the usual astronomical instruments like sun dials, the astrolabe (Figure 1) became a particularly privileged instrument. Some of them are works of art. The astrolabe is a projection of the sky onto a plane. The most popular is the planispheric, where the celestial sphere is projected onto the plane of the equator. A typical astrolabe was made of brass and was between 10 and 20 cm in diameter. It consists of a hollow disk, called the mater, which is deep enough to hold one or more flat plates called tympans, or climates. A tympan is made for a specific latitude and is engraved with the stereographic projection of lines of equal altitude and azimuth representing the portion of the celestial sphere which is above the local horizon. The astrolabe serves in solving problems relating to time and the position of the Sun and stars, and to measure their altitudes. They were introduced to Europe from Islamic Spain at the end of the 10th century, and were in use until about 1650, when they were replaced by more specialized and accurate instruments.

4.3. Theory

4.3.1. *Arabic numbers*

The first great advance on the inherited Greek mathematical tradition was the introduction of "Arabic" numerals, which actually originated in India and which simplified calculations of all sorts and helped the development of algebra. The introduction of this system is attributed to al-Khwarizmi (780-850), who is also an important figure in the invention of algebra.

4.3.2. *Trigonometry and algebra*

To measure angles Ptolemy had used the chord (Figure 2). In India they invented the sinus. The other trigonometric functions and trigonometry itself, including spherical geometry, are important original contributions of Islamic astronomers and mathe-

Fig. 1. The astrolabe serves in solving problems related to time and the location of the Sun and stars. They were introduced to Europe from Islamic Spain at the end of the 10th century, and were in use until about 1650, when they were replaced by more specialized and more accurate instruments. Photo: Institut du monde arabe, Paris / Philippe Maillard.

maticians, the same is true of Algebra. Calculating with the sinus instead of the chord and working with the newly discovered trigonometric relations greatly simplified astronomical research on the celestial sphere.

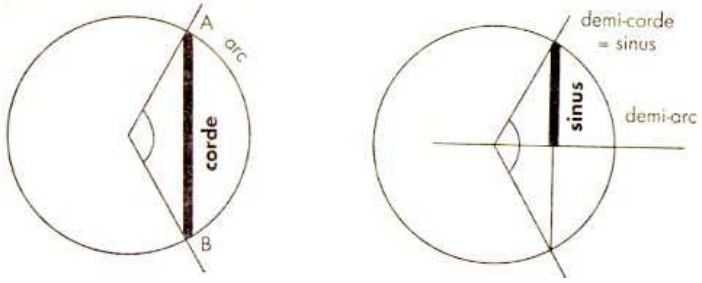


Fig. 2. Chord and Sinus. If the circle has a radius of length 1, then the sinus of half the angle from A to B is half the chord of the angle from A to B.

4.3.3. *Trepidation*
Trepidation is a concept and theory that formed an essential part of medieval astronomy. Copernicus dedicates a whole section of De Revolutionibus to trepidation. Tycho Brahe finally did away with it. What is it? When Islamic astronomers compared their observations with those of Ptolemy, they made two discoveries. The obliquity of the ecliptic was smaller and the precession rate was faster than it had been in Greek antiquity.

Thabit ibn Qurra (836-901) came to the conclusion that this change varied in time. This led to the theory of trepidation. For Ptolemy the 8th sphere, with the ecliptic inscribed, belongs to the fixed stars. Thabit then adds a 9th sphere. It is the sphere of diurnal motion and therefore contains the celestial equator. He then introduced a small circular movement of the 8th within the 9th sphere that would produce what he thought was a cyclic movement. He thought of a period of about 4000 years. From the 12th century onward it was realized that precession was going in only one direction, but that the plane of the ecliptic had a small wobble.

4.3.4. *Criticism on Ptolemy*
Although Ptolemy's Almagest remained the basis of Islamic astronomy, that work was attacked since the 9th century. This culminated in a first major critique by Nasir al-Din al-Tusi (1201-1274) and his students, especially Qutb al-Din al-Shirazi (1236-1311). One of the main points of attack was the equant in the theory of the planets and the moon, because it violated the principle of regular circular motion.

Fig. 3. Retrograde motion of the outer planets. The path of Mars from June to September 2003.

This was not just a technical point. It was an essentially philosophical question about the nature of the universe. According to Aristotle the heavenly bodies are immune to changes in their nature, and there is only one kind of motion fit for them, this is the steady circular motion, because this kind of movement has no beginning and no end, and any position on a circle is as good as any other position on this circle. Thus, Ptolemy had been confronted with the problem of describing the observed varying speed of a planet by a model of constant circular velocity. For this he invented the equant. The centre of the epicycle moves on a circle - the deferent. To better reproduce observations, this circle is not centred on the earth and the centre of the epicycle moves with varying speed along the deferent. However, he found a point, the equant (Figure 4), from which the centre of the epicycle is seen to move with constant angular velocity. But the philosophers saw this as an artifice and inadmissible violation of the Aristotelian principle. Ibn Rushd (1126-1198, latinized

Averroes) criticized eccentric spheres and the epicyclic sphere as contrary to nature. He reproached the Ptolemaic system, that it hold no deeper truth, but was simply a mathematical tool to describe celestial movements.

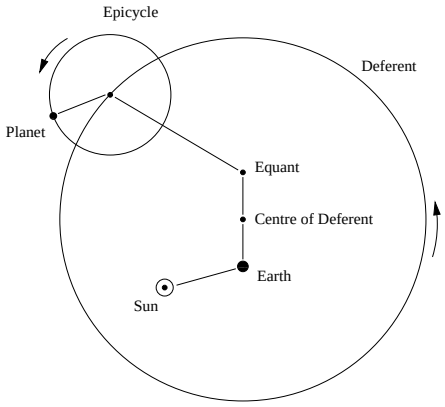


Fig. 4. The Equant. Seen from the Equant, the centre of the epicycle moves with constant angular velocity.

Islamic scientists developed mathematical tools to replace the equant. The most prominent are al-Urdi (d.1266), al-Tusi (1201-1274), and al-Shatir (ca.1304-1375) Saliba (1994, 2005). Al-Tusi and his group discovered, that by combining two uniformly revolving epicycles, they could generate an oscillation along a straight line. Today, this is called the Tusi-couple (Figure 5). They devised a model that produces a linear back and forth motion out of two circular motions, which allowed them to do away with Ptolemy's equant point by adding two additional small epicycles to the excentric planetary orbit. Al-Shatir of Damascus (ca.1304-1375) eliminated the equant as well as the excenter by minor epicycles. This made it possible to explain the non-uniform motion of the epicycle around the deferent in terms of purely uniform motions.

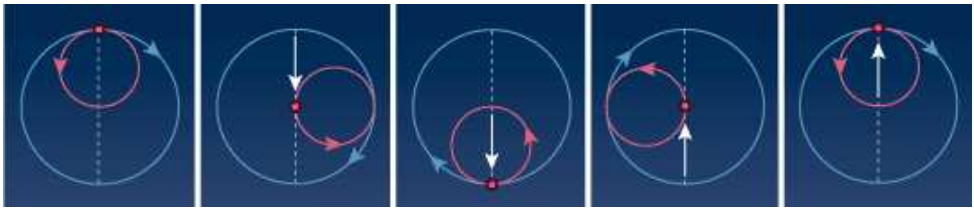


Fig. 5. The Tusi couple. The combination of two steady circular motions results in a back and forth movement along a straight line. Having a circle of half the radius of the epicycle turn tangentially in the epicycle can replace the equant.

But why did Islamic astronomers not challenge the geocentric system? Well, there was no observational evidence against it. Even such a clever observer as Tycho Brahe refused the Copernican idea and developed his own geocentric system. One of the main objections was the absence of yearly stellar parallaxes. The first

observational proof that a planet was circling the sun and not the earth came in 1613, when Galileo Galilei discovered the changing phases of Venus.

4.4. Astrology

Astrology was rejected by both Islamic and Christian religious leaders. On the other hand, philosophers looked to Aristotle and his analogy between the macrocosmos of the universe and the microcosmos of the human body. Ordinary people as well as the caliphs and kings were always much interested in what astrology had to tell them. Astrology is one of the most important historical contexts in which astronomy developed. In the West astronomy and astrology went closely together up to the beginning of the 17th century.

Fig. 6. Albumasar (787-886): Introductorium in astronomiam. Printed 1489 in Augsburg and 1506 in Venice, it proves the influence of Islamic astronomy and astrology on Renaissance Europe. (Zentralbibliothek Zürich)

Important Islamic astrologers had a high social standing. The court astrologer of Harun al-Rashid was supervisor of the royal library, which shows, that these personalities had a profound knowledge of the classical texts available at the time. To the Medieval West the most important representative of Islamic Astrology was the Persian Abu Ma'shar (latinized Albumasar, 787-886). His works (Figure 6) were translated into Latin in 1133 by John of Seville and circulated widely in manuscript form. He exerted a powerful influence on the development of Western Astrology. And we should never forget, that these astrologers were scholars of a wide culture, who wrote on philosophical and astronomical subjects as knowledgeable as about

astrology.

5. What did the Christian culture at the time of Charlemagne know about astronomy?

Very little, is the most concise answer. When during the decline of the Roman Empire the Völkerwanderung upset in Europe the old order of values, the cultural heritage of antiquity was practically lost. Some books were preserved in monasteries, but their content was hardly understood. Religion became the all dominating factor. Charlemagne lived at the same time as Harun al-Rashid; they had political contacts, but the difference in the scientific levels of the two cultures was enormous.

Ambrosius Theodosius Macrobius lived in the early 5th century and wrote a comment about the "Dream of Scipio". This comment contains a section on astronomy. An outermost sphere is heaven itself, the Supreme God, enveloping and comprehending everything in existence, including all the fixed stars. Inside this one lie the spheres of Saturn, Jupiter, Mars, the Sun, Venus and Mercury, the Moon, and innermost the sphere of the Earth. The movement of the spheres creates the heavenly music. However, there is no detailed astronomical information in that work. Of course, the Christian Church needed some astronomy to set the Easter date. This kept the interest in astronomy alive, though on a very low level.

Another, well known source about the heavens was Martianus Capella of Carthage who lived about 365-440. For his son's wedding he wrote an allegory "De nuptiis Philologiae et Mercurii (The Nuptials of Philology and Mercury)". In this encyclopedia on the seven liberal arts we find a treaty on astronomy. It is a qualitative narrative about the geocentric system, without any detailed description or prescription how to find the paths of the planets. However, it contains the interesting statement, that Venus and Mercury circle the sun - Copernicus (1543) mentions that passage. But compared to the Almagest, the astronomical information in Martianus Capella is very primitive. This was the status throughout the first millennium.

The development of cathedral schools in the eleventh and twelfth centuries, as part of the reform program championed by the papacy, brought renewed interest in the heritage of classical antiquity. At Chartres, for example, there was extensive study of Plato's cosmological work, the Timaeus, along with Chalcidius'(4th century) commentary, Martianus Capella, Macrobius, Seneca's "Natural Questions", Cicero's "On the Nature of the Gods", along with works by Augustine, Boethius, and John Scotus Eriugena. Timaeus was particularly important in that it contained the most systematic discussion of questions in cosmology and physics. Thierry of Chartres (d. after 1156) was especially influential in his attempt to use Platonic cosmology in his reading of the creation account in the Genesis. But, schools established at cathedrals and monasteries, as well as at secular courts, were much more preoccupied with grammar, logic, theology, and biblical exegesis, rather than with scientific questions.

6. Contacts with the Islamic culture and translation centres

In the 10th century contacts between Islamic and Christian intellectuals were still rather sporadic, but becoming more frequent. One example is Gerbert of Aurillac

who was born between 940 and 950. He entered the Church, was interested in natural sciences, traveled to Spain where he studied in Ripoll and Vic, and possibly with Islamic teachers in Cordoba and Sevilla. He was interested in astronomy, and learnt the Arabic number system. On his return he became teacher at the school of the cathedral of Reims. It is likely that he brought from Spain the knowledge of the astrolabe to the cathedral school of Reims. He was far superior to anyone else in doing mathematics. Thus, with the astrolabe and with a new system of numbers, he was watched with suspicion, and even suspected of being a magician in touch with the devil. He was later elected Pope Silvester II (999-1003).

The gates really opened after the conquest of Toledo in 1085 by Alfonso VI. The large scale re-conquest had started with Fernando I in the middle of the 11th century. A scholarly community, "The School of Toledo", grew under the leadership of Archbishop Raymond of Toledo who guided that institution from 1126 until his death in 1152. Raymond knew about the wealth of scientific expertise of the Islamic culture. He created a "Translation Centre". The purpose was to translate into Latin all the Greek and Arabic manuscripts. Raymund's centre employed Jewish, Christian, Muslim, Latin, and Greek scholars.

The centre attracted renowned thinkers from all over Europe such as Robert of Chester, Adelard of Bath, and David Morley from England; John of Brescia, Plato of Tivoli, and Gerard of Cremona from Northern Italy. Every available work on astronomy and astrology was translated. Robert of Chester traveled from England to Toledo to learn Arabic, and in 1145 completed a translation of al-Khwarizimi's treatise on calculation. Around 1175 Gerard of Cremona came to Toledo to translate the Almagest from Arabic into Latin. This version was copied by hand, until in 1515 it was printed in Venice. It was only at that time that translations were made directly from the Greek text which had come to Rome after the Turks had conquered Constantinople in 1453. After the death of Raymond in 1152 the school of Toledo continued into the 13th century. It worked for a while in parallel with the new translation centre of Alfonso X (1221-1284), called "el Sabio" (the Learned).

Two astronomical works gained particular importance: The Toledan Tables and the Alfonsine Tables. They give the recipe for determining the astrologically all important positions of sun, moon and planets. Azarquiel or Arzachel (al-Zarkali 1028-1087) worked in Toledo and produced the Toledan Tables in Arabic. Already in 1140 they were translated by Raymond of Marseille and adapted in the "liber cursuum planetarum" to the meridian of Marseille. They had a wide circulation because they allowed to calculate planetary positions for any time. However, in order to really work with them, an understanding of the Almagest was needed. At the request of Alfonso X (1221-1284) the Toledan Tables were translated around 1255 into Castilian and again around 1277.

The best known and influential work through the later Middle Ages are the Alfonsine Tables (Figure 7). Alfonso X had inherited a vastly extended kingdom. In addition to Toledo it included Cordoba and Seville. This gave the Christian scholars access to the Andalusian culture that had been flourishing during the previous centuries. Happily the kings of the re-conquista, Alfonso VI, Ferdinand III, and Alfonso X were similarly open minded as the Califs of Baghdad in the 8th and 9th century. They showed great interest in culture and a high regard for what they found

in their newly conquered places. Alfonso X gathered around him a large group of scientists, mainly astronomers, who translated from Arabic mainly into Castilian, but occasionally also into French and Latin.

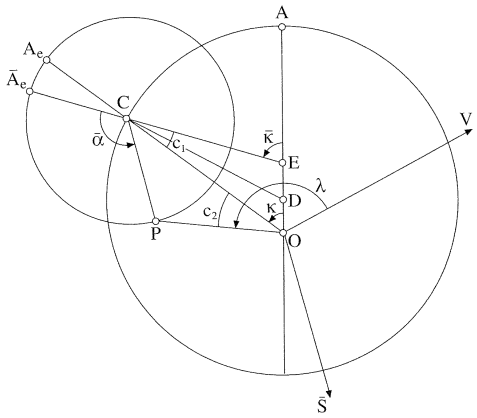


Fig. 7. The Alfonsine Tables of Toledo (Chabas & Goldstein, 2003). The Tables give procedures to calculate the longitude of the sun and the planets on the ecliptic.

However, they not only translated, they also created original works. The Alfonsine Tables are such a creation. They are not the work of a single author, but of a whole group. The members of the group mastered more than one language and their members had different cultural backgrounds being Christians or Jews or coming from Islam. The Tables contain the technique of trigonometry, daily rotation, latitudes, planetary visibility and retrogradation, planetary and lunar velocities, Syzygies (three bodies found along a straight line) and eclipses, visibility of the lunar crescent, astrology, mean motion, the equation of time which gives the deviations from the mean motion, trepidation. The Alfonsine Table take as starting point the year 1252, 1252 was the coronation date of Alfonso X.

The book on stars of al-Sufi had a lasting influence on stellar toponymy in European languages. It was translated into Spanish by Alfonso X el Sabio. Many of our star names, such as Aldebaran, still recall their Arabic origin.

The intellectuals gathered in Toledo gained access to new ways of thinking, and were able to incorporate into their own writing what they found in the Islamic culture. Thus, Adelard of Bath wrote his "Questiones Naturales" as a summary of ideas he collected from Arab sources. Adelard, who translated al-Khwarizmi's Astronomical Tables and "Liber Ysagogarum" (About Arithmetic), later wrote books such as the "Rule of the Abacus" and the "Usage of the Astrolabium", which were strongly influenced by al-Kwharizmi's teachings.

7. The impact of Islamic astronomy

European philosophy, science and medicine received an enormous impetus from the contact with the Islamic world in all branches of culture: Medicine, philosophy,

literature, music, mathematics, astrology, astronomy, geography and others. We concentrate on astronomy.

The Arabic numbers came to Europe through several channels. The first one was probably Gerbert of Aurillac (ca. 940-1003), whom we already mentioned. He studied in Islamic Spain and later became Pope Silvester II. Then we have the 12th century translation al-Khwarizmi by Adelard of Bath. The use of Arabic numbers became more widespread when the Italian Fibonacci (ca. 1170-1240) published his "Liber Abaci" in 1202. Fibonacci had lived near Algiers for several years, where his father was ambassador for Pisa. His book was meant to help commerce. But Roman numbers were used up to the time of Georg Peuerbach (1423-1461), who wrote the first German book on calculations, using Arabic instead of Roman numbers.

Islamic culture did not simply hand over what it had inherited from Ptolemy. They created new mathematics, trigonometry and algebra, they improved the Ptolemaic methods and tables, and thoroughly criticized Ptolemy on philosophical grounds.

Many Islamic authors were translated into Latin and influenced European thinking. One of them is al-Farghani who between 830 and 860 wrote "Elements of Astronomy" which is a non-mathematical summary of Ptolemaic astronomy. In Toledo it was translated into Latin first by John of Seville in the first half of the 12th century, and later by Gerard of Cremona (ca. 1114-1187). From this source Dante learnt about Ptolemaic astronomy for the "Divina Commedia", where the poet ascends through the spheres. From the translation of John of Seville Sacrobosco got his information about the Ptolemaic universe. Sacrobosco (d. 1256) is famous for his textbook "De Sphaera" (ca. 1230) which was a watered down edition from al-Farghani, explaining spherical astronomy. According to Gingerich (1986) this book went through 200 editions before it was superseded in the early 17th century.

Al-Battani (850-929, latinized Albategnius) wrote with "Kitab al-Zij" one of the early important works. It contains 57 chapters and begins with a description of the division of the celestial sphere into the signs of the zodiac and into degrees. The necessary mathematical tools are then introduced such as the arithmetical operations on sexagesimal fractions and the trigonometric functions. Chapter 4 contains data from al-Battani's own observations. Chapters 5 to 26 discuss a large number of different astronomical problems, some already treated in the Almagest. The motions of the sun, moon and five planets are discussed in chapters 27 to 31, where the theory given is that of Ptolemy. Al-Battani covers in Chapters 49 to 55 astrological problems, while chapter 56 discusses the construction of a sundial and the final chapter discusses the construction of a number of astronomical instruments. The "Kitab al-Zij" was translated into Latin as "De motu stellarum" (On the motion of the stars) by Plato of Tivoli in 1116 while a printed edition of this translation appeared in 1537 and then again in 1645.

Al-Battani's work was very influential in the West. Tycho Brahe knew it, and Copernicus mentions him. Let us come back to trepidation and equant. Tables of trepidation were included in the Toledan Tables, finalized by al-Zarqali in about 1080. But then it was realized, that precession was not periodic, but continuous. The makers of the Alfonsine Tables introduced a steady precession with a period of 49'000 years and an oscillation of +/- 9 degrees which is completed in 7'000 years.

This became the standard theory from the 14th to the end of the 16th century. But this steady precession needed an additional 10th sphere. The 8th sphere is the firmament, the 9th sphere is occasionally (Apian) called crystalline, it is responsible for the trepidation of the 8th. The 10th sphere is called the first movable, it gives the daily rotation and is responsible for precession of the 9th sphere. Copernicus blamed the periodic component of precession and trepidation on the movement of the Earth's axis. However, Tycho Brahe came to the conclusion, that precession was a steady process and that trepidation is probably a periodic wobble of the plane of the ecliptic.

In the Christian Middle Ages, particularly during the scholastic period, the heavenly spheres were the subject of much speculation. The discussion was split into subjects of astronomical and religious relevance. Dante (1265-1321) in the "Divina Commedia" is led by Beatrice through the nine spheres of Heaven. They are: (1) The Moon, (2) Mercury, (3) Venus, (4) the sun, (5) Mars, (6) Jupiter, (7) Saturn, (8) the stars, (9) The Primum Mobile (the Prime Mover).

Copernicus was as unhappy with the equant as al-Tusi (1201-1274) and al-Shatir (1304-1375) had been before him. Copernicus (1514) wrote in the Commentariolus: "I often pondered whether perhaps a more reasonable model composed of circles could be found, from which every apparent irregularity would follow, while everything in itself moved uniformly, just as the principle of perfect motion requires." This same problem had been discussed and solved by al-Tusi, who showed that a combination of two circular motions can result in a straight line. Copernicus (1543) employed that procedure in De Revolutionibus, Chapter 4 of Book 3 which carries the title "How an oscillating motion in libration is constructed out of circular motion". There he tried to explain trepidation, the non-uniformity of precession. He explained the backward and forward motion by the Tusi couple, however, he does not mention al-Tusi. He also replaced the equant with epicycles in the way al-Shatir had done before him. It is very likely that the mathematical models he employed were inspired by these two scientists. (See also Swerdlow 1996, and Dobrzycki & Kremer 1996).

It is not known how Copernicus learnt about the models of al-Tusi and al-Shatir. But, because of the many similarities, historians have for some time been quite certain, that he must have known those models. He may have become acquainted with them through a 14th century manuscript of Ibn al-Shatir containing the Tusi couple, which was a translation of an Arabic text into Greek, made probably in Constantinople. It might have reached Rome after the fall of Constantinople in 1453, and is still in the Vatican library. Copernicus may have seen it during his sojourn in Rome in 1500 (Saliba, 1999).

8. Conclusions

The influence of the Islamic culture on Europe from the end of the 10th to the 13th century was very great in all domains, in poetry, music, philosophy, medicine, and science. Let us first summarize the conditions that favoured the emergence of the Islamic culture:

1) There was a strong economical basis. Conquests between 635 and 750 had resulted

in an immense empire, controlling essential resources and trade routes from the Indian Ocean to the Atlantic.
2) Arabic was the lingua franca for the entire cultural room. The arrival from China of the art of manufacturing paper at the time of Harun al-Rashid gave an enormous boost to the dissemination of knowledge.
3) An open mind for the values of other cultures. A spirit of religious tolerance allowed Muslims, Jews, and Christians to work side by side.
4) Education was wide-spread and touched all layers of society.

Astronomy had a very high social standing. It was very much an applied science. It delivered calendars, the techniques for exact geography, and it dealt with the high art of astrology. Islamic astronomy further developed the techniques inherited from Ptolemy.

Islamic astronomers did more than simply read Ptolemy. They provided astronomy with new mathematics. Their invention of algebra and trigonometry substantially changed the mathematical foundation of astronomy. They also questioned Ptolemy fundamentally. They created cosmological alternatives. However, they never questioned geocentricity.

Islamic astronomers were well known in the Christian Middle Ages. Copernicus (1543) mentions some in his De Revolutionibus, e.g. Ibn Rushd, al-Battani (his name I found 27 times), Al-Zarkali (mentioned 8 times), Thabit ibn Qurra, and al-Bitruji (latinized Alpetragius, died 1204 in Spain). The fundamental criticism on the Almagest by al-Tusi and others may have been instrumental in the work of Copernicus.

The revival of European learning which began slowly at the end of the 10th century, was strongly influenced by contacts with the Islamic world. These contacts, first a trickle and eventually a flood, radically altered the intellectual life of the West. They familiarized Europeans with the positional numerals and the zero. They acquainted European astronomers with Ptolemy's Almagest. Its translation from Arabic into Latin in the 13. century, combined with the additional work of the Islamic scientists brought scientific astronomy to Europe. It familiarized Europeans with scientific thinking. It gave in the 13. century the initial impetus to European astronomy.

References

Chabas, J., Goldstein, B.R.: 2003, The Alfonsine Tables of Toledo. Kluwer Dordrecht.
Copernicus, N.: 1514, Commentariolus. English Translation by N.M. Swerdlow in Proc. Amer. Phil. Soc.**117**, 435.
Copernicus, N.: 1543, De Revolutionibus Orbium Coelestium. English Translation by Edward Rosen. The Warnock Library.
Dobrzycki, J., Kremer, R.L.: 1996, J. for the History of Astron.**27**, 187
Kennedy, H.: 2005, When Baghdad ruled the Muslim world: the rise and fall of Islam's greatest dynasty. Cambridge, Mass.: Da Capo Press, Dordrecht, Kluwer
Gingerich, O.: 1986, Scientific American **254**, 74.
Gutas, D.: 1998, Greek Thought, Arabic Culture. Routledge, London and New York.
Saliba, G.: 1994, A History of Arabic Astronomy. New York University Press.
Saliba, G.: 1999, Whose science is Arabic Science in Renaissance Europe?, Columbia University, Internet.
Saliba, G.: 2005, L'astronomie arabe. L'age d'or des sciences arabes. Institut du monde arabe,

Paris. Catalogue de l'exposition, p.53-67.
Swerdlow, N.M.: 1996, "Astronomy in the Renaissance" in Astronomy before the Telescope, ed. Ch. Walker, British Museum Press.

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

ASTRONOMY AT MESOPOTAMIAN REGION (3000BC-1400AC)

HAMID M.K. AL-NAIMIY
College of Arts and Sciences, University of Sharjah, Sharjah, UAE

Abstract. I - The Babylonian (Mesopotamian BC) astronomy achievements have been divided into three classes according to the type of activity and period, such as:

1. The early schematic Babylonian astronomy for the period (3000 - 750 BC), including :

- Observing the periodical and astronomical phenomena such as Eclipses.

- Simple schematic and linear equations for celestial phenomena.

- Astrolabe texts and the helical stars rising along the eastern horizon.

- The calculation of the duration of days and nights throughout the year, produces a function known as the linear ZIG-ZAG function.

2. The Non-Mathematical astronomy (depends on watching the sky and the omens) during the period (750 - 350 BC), including: Astronomical Diaries, Almanacs and the Goal year text.

3. The Mathematical astronomy from the period (350 - 50 BC), including: The astronomical Ephemeris or tables of lunar, planets and stars phenomena, with other important achievements computed at a regular interval, and astronomical observatories. Besides, the oldest mathematical linear function appeared from the Babylonian astronomical measurements from uniform equations of the relation between lunar and synodic year. At the same time we should mention the discoveries of the 12 zodiacal constellations, the circular astrolabes, the Almanacs, the Ephemeris and other astronomical issues, including, the tables of observation or prediction of the Lunar and Solar Eclipses.

II - The paper contains as well the most important achievements of the Arab and Muslim Astronomers in Astronomy in the same region during the period 900 - 1400AC.

1. Introduction

It is due to the clear sky that is free from any lights and Industrial pollution that the Mesopotamians became the pioneers in Astronomy. Also the beauty of their lands and the fertility and the availability of water on them, accelerated the emergence of the first human settlement especially on the sediment level land from the southern of Iraq (down the Euphrates), the ancient Iraq, or the land between the rivers Tigris and Euphrates.

In addition, the shortage of their lands to some raw materials like woods, valuable stones and metals made them import these materials from the neighboring countries, therefore, they became warriors, tradesmen, farmers, and they prepared the armies to protect the transport ways and they went on the deserts and seas. So, they resorted to the use of the stars to know the routes during the nights, because without the stars the armies would loss their ways and their caravans would be in the great

danger of the deserts and the seas. By the year 3000 BC, the Mesopotamian culture had developed an irrigation system, building methods using clay bricks rather than wood or mud, and a system of writing. For Example: (Ziggurats: the Mesopotamians built of clay bricks on the flat flood plain of the Tigris and Euphrates rivers), (Ahmad, 2003).

They also resorted to the use of stars in order to set the calendars determining the times of sowing and harvest, flood, and the feasts. They also used the phases of the moon, zodiac of the sun and the positions of the planets to predict the states of the weather and to assign the dimensions of the countries and the calculation of the crescent appearance and other matters that are connected with their daily life (Al-Naimiy, 1990).

Astronomy had been a very interesting subject that stirred the interests of people in the former centuries. As such various names have been given to this science. Some of these are really the science of stars, the science of the shape, the science of the universes, the science of astrology, the science of creating stars and astronomy, these are the commonest and the most famous names in our legacy. But the name astrology is the strongest due to its deep roots in our culture. It dates back in its origin to the Babylonian work "Pulukku" (Al-Naimiy, 1990), where it means the monitoring of stars, and if we look a little bit closer in the derivation of term "astronomy" and "astrology", we would find that they originate from the Latin term "Astro", which means star and refers to the planets and the universes in general. This Latin word is taken from the name of the Babylonian goddess "Ishtar" that was shifted in the Hebrew language into "Aster" then into "Astr". The evidence is that there were goddess for justice in Greece called "Astraea" which is derived from the Babylonian "Ishtar" as well. It is known that the Babylonian "Ishtar" and before it the samurai "Inana" represents planet Venus, the goddess that was very popular at that time as morning star, evening star and it was a symbol for beauty and loyalty (Rashid, 1987). It was also believed that its emergence and disappearance impact the life of the Man on Earth. While the Greeks term was "Astronomia" which means in Arabic the science of stars and its literal translation "Astr" means star while "Nomia" means science. In fact, the beginning of the predictable experimental astronomy started in Mesopotamia throughout their relation with the agricultural purposes that are connected with the rain. This started approximately 3000 years BC when Astrology was activated with the growth of magic as well. This paved the way to evolution of clear religion later on, and this goes in line with the idea that astronomy or astrology was started basically to establish an earthly and celestial religion (Ahmad, 2003). Therefore, in this research we will not come across the Babylonian astronomy before the period 3000 years BC, while we are going to look for the Babylonian astronomical achievements during the period 3000 BC till 50 A.D. This was the beginning of real astronomy according to our estimation that involves the Sumerian, Caledonian, and the Babylonian Astronomy. We are going however to call it generally the Babylonian Astronomy in order to agree with the common European names nowadays. We will try to follow a division to the achievable and the chronological order.

What I am after in this regard is really the documented astronomy in terms of the observations, predictions, and the calculations, in addition to the practices of the celestial anticipation in the ancient Mesopotamia with the intention of stopping by

the basic differences that discriminate these practices with their alternatives in the Greece countries during the same period or after on one hand, and what is available right now on the other hand. Although there was a big deal of information still unknown so far and so many questions need to be set. In this paper we are speaking about the ancient Iraq or the land between Tigris and Euphrates. In this area we find that the evolution of the precise sciences such as Mathematics and astronomy, were developed alone without other sciences in the sense that in some civilizations, for instance, astronomy had never reached the status of a scientific subject due to the primitivism of the mathematical tools. Therefore, this science had stayed in the preliminary informative and the planning level; also it hasn't any influence on the other cultures. While in Mesopotamia astronomy, on the contrary from, had been accorded a dignified status in accuracy and its impact had been stretched to the futuristic developments in the Greek and Indian sciences and later in the Islamic ones. The Babylonian astronomy used to have a very influential impact on the Greek world due to the availability of the great excess of amount for understanding the celestial phenomena. In spite of the co-occurrence of the last two periods from the Caledonian astronomy with the Greek astronomy, the Caledonian one was the first source for the celestial information and tables. The Caledonian excellence in astronomy was very evident even in its arousing from the first philosophical Greek questions. On this basis, we have been able to by the help of Allah divide the Babylonian astronomical and intellectual achievements during the period from 2500 BC to 50 BC into three divisions according to a chronological order that fit with the astronomical achievement and the text at the hand. These divisions are as follow:

1. **Early Babylonian Astronomy:** it included the period 2500BC - 750BC with many achievements such as
   - Control of the horizontal and the periodical phenomena for the heavenly bodies.
   - Schematic numerical Functions and Linear methods for heavenly phenomena.
   - Astrolabe texts and the Plough text (number of the big bear group and the routes of the goddess in the eastern side of the sky) then the motion of the planets and stars in those routes.
   - The calculations that involved the Zigzag function (triangle of day and night calculations hours by the use of the water watch with the passage of the year) and assigning the positions of the Fall Equinox and the Spring Equinox and the Summer Solstice and the Winter Solstice, then the calculation of Al Saros cycle that ranges 2/3 18 years.
2. **Non Mathematical Astronomy:** (during the period 750-350BC) it included special astronomical calendars like the daily diaries and the (almanacs) and texts for a certain year called goal year.
3. **Mathematical astronomy:** (from 350-50 BC) it included the important tables that are called "Ephemeris" and it also included lunar and planet Ephemeris, besides recording the movements of the planets near the zodiacal area and other astronomical subjects that are very important and which based on mathematical calculations namely those that lasted for long centuries after the year 50 BC.

In the next sections follow the important Babylonian astronomical achievements after they have been shifted into an astronomical scientific and mathematical lan-

guage.

2. The Mesopotamian Achievements (3000 - 50 BC)

- The Babylonian approach to astronomy depended on two basic methods: Observation and Computation. Both are found in the earliest cuneiform texts (i.e., dated to the Old Babylonian Period).
- The main sources of the Babylonians from about 1800 BC to about 500 BC are the **Ea, Anu and Enlil** omen series, the circular and tabular "astrolabes" (i.e., planispheres), the MUL.APIN series, and various observational texts (i.e., reports to the Kings and the earliest astronomical diaries).
- From about 1800 BC to about 500 BC the main phenomena the Babylonians sought to be proficient with were:
  (1) The appearance and disappearance of Venus;
  (2) The duration of day and night;
  (3) The rising and setting of the moon;
  (4) Planetary and stellar risings and settings. All appear within the proteases of the celestial omens of the 2^{nd} millennium period (i.e., the Ea, Anu and Enlil series).

3. The Sumerian and Akkadian Period (3200-2000 BC)

During this period the Sumerians watched the sky and defined and named some of the constellations and planets. Most of the names of celestial bodies were Sumerian throughout the later periods and some of them at least must have Sumerian origins.

4. Babylonian, Assyrian and Chaldeanian era (2000 - 50 BC)

I) The old Babylonian Period In this period the recorded phenomena were:
- (1) duration of day and night (before and after 750BC);
- (2) rising and setting of the moon;
- (3) Appearance and disappearance of Venus.

II) The Period (1500-750BC)
- The composition of the great Omen Series "Ea, Anu and Enlil."
- Exact observations of the heliacal risings of fixed stars.
- Observations of daily risings, culminations, and settings.
- Composition of the circular and rectangular Astrolabes before 750 BC.
- A very primitive representation of the Venus phenomena by arithmetical sequences.
- Calculations of the lengths of day and night by increasing and decreasing arithmetical series

III) The Late Assyrian Period (750-600 BC) The systematic observation of celestial phenomena began in the Assyrian Period and continued without a break into late Seleucid times (Astronomy of the MUL.APIN series).

The main astronomical achievements of this period were:

– Detailed study of the fixed stars, their risings, culminations, and settings.
– Calculations of the duration of daylight and the rising and setting of the moon by "linear methods".
– Recognition of the zodiac as path of the Moon, the Sun, and the planets.
– Establishment of zodiacal constellations.
– Position of the zodiac with regard to the zones of Enlil, Anu, and Ea.
– Establishment of the seasons of the year.
– By about 750 BC the calendar became astronomically regulated by the risings of stars and constellations.

IV) The Chaldean Period (600-500 BC) In this period the main astronomical features were:
– Progress towards the division of the zodiac into 12 signs of 30 degrees each.
– Systematic observation of the Moon and the planets, their positions in relation to the fixed stars, their first and last visibility, stationary points, conjunctions, etc.

V) Neo-Babylonian, Persian and Seleucid Periods (500-50 BC)
– Mathematical astronomy. The largest and most highly developed part of the theoretical astronomy of the Seleucid period is devoted to the computation of the new moons.

The astronomy of the Neo-Babylonian and Persian period has the following typical features, (Ahmad, 2003):
– Systematic, dated and recorded observations of eclipses and lunar and planetary phenomena.
– Calculation of Periods.
– Prediction of eclipses.
– Division of the zodiac into 12 signs of 30 degrees each.
– Rise of horoscope astrology.
– Development of mathematical astronomy.

The most important achievements of this period are:
– Determination of accurate periods for the Sun, the Moon, and the planets.
– Calculation of the motion of the Sun, the Moon, and the planets, of eclipse magnitudes and other lunar and planetary phenomena. (These calculations were based upon an admirable mathematical theory.)

5. Babylon and Sumerian Records of Eclipses

1. Babylonian clay tablets that have survived since dawn of civilization in the Mesopotamian region record the earliest total solar eclipse seen in Ugarit on May 3, 1375 BC. Like the Chinese, Babylonian astrologers kept careful records about celestial phenomena, including the motions of Mercury, Venus, the Sun, and the Moon on tablets dating from 1700 to 1681 BC.
2. Later records identified a total solar eclipse on July 31, 1063 BC, that "turned day into night," and the famous eclipse of June 15, 763 BC, recorded by Assyrian observers in Nineveh. Babylonian astronomers are credited with having

discovered the 223-month period for lunar eclipses. One of the famous total solar eclipse recorded was the one on 15^{th} of April, 136BC. Many stars and 4 planets were seen (Stephenson, 1982)

3. The regular & recorded observed lunar phenomena:
   a) Observed just after New Moon on the evening of first visibility of the crescent time between setting of sun and moon on the evening of the first visibility of the crescent.
   b) Observed just before and after Full Moon.
      – Time between the last setting of the moon before sunrise and sunrise.
      – Time between the last rising of the moon before sunset and sunset.
      – Time between sunrise and the first setting of the moon after sunrise.
      – Time between sunset and the first rising of the moon after sunset.
   c) Observed on the day of last visibility of the moon in the morning.
      Time between the rising of the moon and sunrise on the morning of last visibility of the moon just before New Moon.

6. The most important achievements of Babylonian Astronomers

- Saros : 223 Synodic months = 242 Nodical months (18yr + 10.3 day eclipse period).
- Lunar disk divided into 4 sectors ; S = Akkad , N = Subartu , E = Elam, W = Amurru
- The discovery of Linear Zig-Zag function.
- The discovery of Step Function (Width of the 12 Zodiacs: 30^o).
- Prediction of the Lunar & Solar Eclipses.
- Astronomical Diaries, Almanacs, eclipse reports, goal-year texts and Ephemeris.
- The use of the 7^{th}, 4^{th}, 12^{th} and SEXAGESIMAL Systems.
- With the exception of the Venus tables of Ammiza-duga, which probably originated in the seventeenth century BC, most of the surviving Mesopotamian astronomical texts were written between 650 and 50 BC. These clay tablets with cuneiform writing are called astronomical diaries, and they are the unmistakable observations of specialists: professional astronomer-scribes.
- A typical diary entry begins with a statement on the length of the previous month. It might have been 29 or 30 days. Then, the present month's first observation - the time between sunset and moonset on the day of the first waxing crescent - is given, followed by similar information on the times between moonsets and sunrises and between moonrises and sunsets, at full moon. At the end of the month, the interval between the rising of the last waning crescent moon and sunrise is recorded.
- When a lunar or solar eclipse took place, its date, time, and duration were noted along with the planets visible, the star that was culminating, and the prevailing wind at the time of the eclipse. Significant points in the various planetary cycles were all tabulated, and the dates of the solstices, equinoxes, and significant appearances of Sirius were provided.

- The Babylonian astronomers used a set of 30 stars as references for celestial position, and their astronomical diaries detailed the locations of the moon and planets with respect to the stars. Reports of bad weather or unusual atmospheric phenomena - like rainbows and haloes - found their way into the diaries, too. Finally, various events of local importance (fires, thefts, and conquests), the amount of rise or fall in the river at Babylon, and the quantity of various commodities that could be purchased for one silver shekel filled out the diligent astronomer's report.
- By the sixth century BC, Neo-Babylonian astronomers were computing in advance the expected time intervals between moonrise or moonset and sunrise or sunset for various days in the months ahead. These calculations were based on systematic observations. Later, when combined with numerical tabulations of the monthly movement of the sun, the position of sun and moon at new moon, the length of daylight, half the length of night, an eclipse warning index, the rate of the moon's daily motion through the stars, and other related information, these computations enabled reasonably detailed and accurate predictions of what the moon would do and when it would do it (Ahmad, 2003).

7. Stars of Gods Paths

Enlil Path: it contains 33 stars including star groups (Cassiopeia, Orion, Auriga, Cancer, Lion, Corona Borealis, Big dipper or Ursa Major, Dragon, Ursa, Hydrus, Alnaser, Vega, Andromeda as well as Jupiter.
Anu Path: it contains 23 stars including (Aries, Pisces, Pleiades, Scera, Virgo, Canis Major, Algorabe, Lira, Al-Nekab and Al-Dewan).
Ea Path: it contains 15 stars including (Pisces Austrians, Aquarius, Centaurs, Lapus, Scorpios and Bootes.

8. Astronomy in the Islamic era (700 - 1400AD)

- Islamic astronomy became the western world's powerhouse of scientific research during the 9th and 11th centuries AD., while the Dark Ages engulfed much of the rest of the western world.
- The works by Ptolemy, Plato, and Aristotle were translated, amplified upon and spread throughout the Muslim world.
- **Al-Khwarazmi** developed the first table's trigonometric functions, c. 825 A.D., which remained the standard reference well into the modern era. Al-Khwarazmi was known to the west as "Algorism" and this is, in fact, the origin of the term 'algorithm'. Al-Khwarazmi's calculations were good to five places, allowing for unprecedented precision in astronomy and other sciences (Al-Moemeni, 1992).
- **Al-Battani**, c. 850 A.D., began with Ptolemy's works and recalculated the precession of the equinoxes (Nallino, 1969), and produced new, more precise astronomical tables. Following a steady series of advances in Islamic trigonometry, observations by **Ibn Yunus** of lunar and solar eclipses were recorded in

Cairo, c. 1000 A.D. Ibn Yunus is regarded as one of the greatest observational astronomers of his time (Al-Naimiy, 1993).

– The pace of Islamic science and scholarship eventually slowed down in the 14th and 15th centuries. Many great books and great ideas of the Islamic Age lay fallow for hundreds of years until they were finally translated into Latin and fueled the European revolution in thinking and the birth of science as we know it today
– We have made statistics for the Muslims Scientist's contributions (Published Books, Encyclopedias, Ephemerides(Al-AZYAJ)...etc. during the era 7^{th} up to 14^{th} centuries, we found that the highest numbers of contributions were during the 9^{th} and 10^{th} centuries (Al-Naimiy, 1993).
– The Muslim's astronomical contributions in comparison with others have been tabulated in Al-Naimiy (1993).

9. Summary and Conclusion

Babylonian astronomical achievements during 2500-50BC

– SEXAGESIMAL: it was used in calculation, weighing, time measurement, length and width lines and triangles.
– Astronomical calendars and tables, which were derived from calculation and observations (for religious, world and astronomical sakes), a model for Jewish, Greek and Roman calendars.
– Weather change notices as a function of Moon position or Sun position in the signs of Zodiac.
– Dividing signs of Zodiac belt to 12 signs, Moon phases to 28 as well as the appearance of the principle of apparent Sun movement in the celestial globe, which are the principles determining the eclipse period.
– Naming of some stars and constellations, which could be seen by mere eye, especially signs, with animal or Kings or Gods names; some of them were based on certain myths.
– Name of twelve months, which are used in the Arab East.
– Week base according to the four Moon forms: they have had a special importance for the beginning of the month, seventh day, fifteenth day and the end of the month, where festivals and celebrations were made, victims were slaughtered, and washing was done. Sometimes Kings attended to do some business.
– Use of the lunar calendar.
– The discovery of leap rule: adding a full month every three lunar years in order to make the lunar and solar years identical, then leaping 7 times every 19 years, a day every three lunar years to make the year 355 days. As well as the famous Babylonian equation (235 months = 19 solar years).
– Calculation of the solar year (365 days and 6 hours and 50 minutes and 15 seconds), within an error of ± 6.5 minutes.
– The discovery of Saros round (18.6 years), the period between two sequenced eclipses.
– Observing planets in order to make prediction about weather. Observing Venus in order to know fortune and to fix calendars and dates of religious festivals and

to make prediction about weather. They knew the period of its conjunction (584 days) and its appearance five times at the same place every eight years.

- Dividing the day into 6 equal parts, 3 for the night and 3 for the day, each part= 2 Beru, Beru = doubled hour, day = 12 doubled hours divided into 360 Uses.
- There is a special meaning for the number 7 in their daily life (sky=7 levels, week=7, planets=7, winds=7)
- Picture and engraves, in the cylinder seals, contained balls and circles rounded by balls or smaller circles rounded by moons. This may indicate that they have recognized the theory of centrality of the Sun.
- The discovery of the first linear **zigzag** functions describing the difference in the day duration. They used this function for the lunar and astral phenomena in the astronomical texts as well.

Lastly, I want to indicate that, the Babylonian astronomical achievements described here may be a small part of the total: there are hundred thousands of unread mud inscriptions as well as damaged ones. They are all great achievements, and most of them are still used in the present days.

References

Ahmad I.: 2003, Astronomy of Babylon, http://www.angelfire.com/wizard/regulus_antares/astronomy_of_babylon.htm/

Al-Naimiy, H.M.K.: 1990, The Arab Astronomical achievements. Proceeding of the Arab Science Heritage in Basic Sciences. College of Science/Al-Fateh University, Libya.

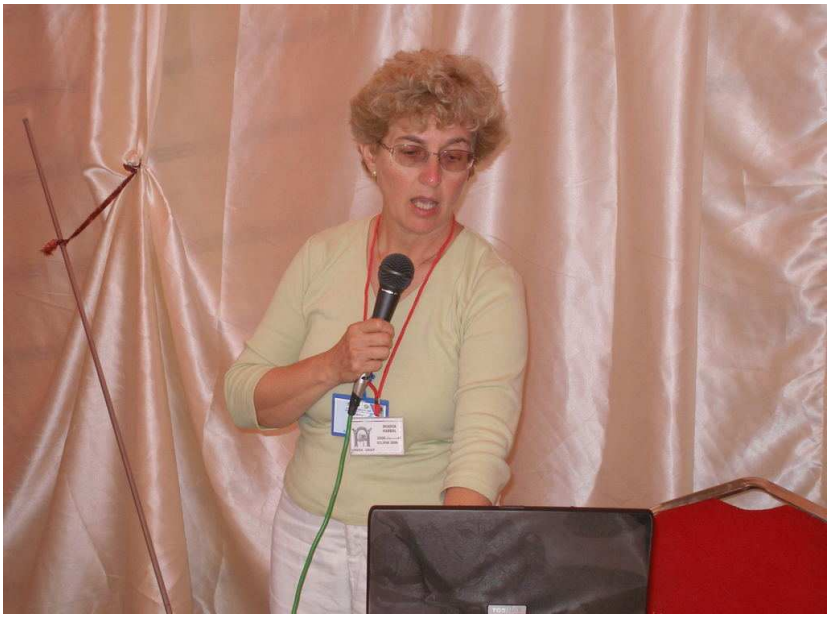
Al-Naimiy, H.M.K. and Al-Waseti, N.L.: 1993, Glories of Arab Thought in Astronomy and Physics, publication of cultural affaires, Iraqi Ministry of culture.

Al-Moemeni, A.: 1992 The Arab and Muslims Astronomical Heritage, Aleppo University, History and Arab Science Institute.

Nallino,C.A.: 1969, Al-Battani sive Albatenii opus astronomicum, Pubblicazioni del reale Osservatorio di Brera in Milano, XL, 3 vols., Milan and Roma; reprinted Frankfurt: Minerva G.m.b.H.

Rashid, F.: 1987, Babylonian Favorites into Astronomy, Arab Astronomical Studies(I), Baghdad University, Iraq.

Stephenson, Richard: 1982, Scientific American, **247**, 154



International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

PROSPECTS OF RENEWABLE ENERGY IN LIBYA

IBRAHIM M. SALEH
Faculty of Engineering, Al-Fateh University P.O. Box 13656 Tripoli, Libya

Abstract. Although Renewable energy applications in Libya were started in the middle of the seventies, they have only gained momentum in the last ten years. Considering the past gained experience, a proposed national Renewable Energy (RE) plan aims toward bringing RE into the main stream of the national energy supply system with a target contribution of 10% of the electricity demand by the year 2020. The proposed plan calls for a wide spectrum of renewable energy applications. This paper will highlight renewable energy applications in the country, the gained experience, the RE resources, and the future prospects for the utilization of RE recourses.

Key words: Stand alone PV systems, Applications and loads, Renewable energy

1. Introduction

Libya is an oil exporting country located in the middle of North Africa, with 6 million inhabitants distributed over an area of 1,750,000 Km^2. The daily average of solar radiation on a horizontal plane is 7.1 kWh/m^2/day in the coastal region, and 8.1 kWh/m^2/day in the southern region, with average sun duration of more than 3500 hours per year (Saleh Ibrahim, 1993). The national electric grid consists of a high voltage network of about 12,000 km, a medium voltage network of about 12,500 km and 7,000 km of low voltage network. The installed capacity is 5600 MW with a peak Load of 3650 MW, for the year 2004 (Saleh Ibrahim *et al.*, 1998). In spite of that; there are many villages and remote areas located far away from these networks. Economically these areas cannot be connected to the grid, owing to its small population, and small amount of energy required. In the past these facts dictate the use of diesel generators as a power supply. The use of diesel generators needs continuous maintenance, continuous supply of fuel. For these reasons we are pushed to look into some other sources like renewable energy. Moreover renewable energy provides clean and reliable energy sources which can be used in many applications in remote areas (electricity, water pumping, etc.). The use of renewable energies has been introduced in a wide range of applications due to its convenient use and being economically attractive in many applications. The most important renewable energy sources are solar energy, wind energy, and biomass.

Photovoltaic conversion, which is the direct conversion of solar energy into electricity, may be considered as the most reliable source for rural electrification. The use of wind energy to electrify remote areas will not be a reliable source as wind energy does not guarantee a continuous supply. Beside that the use of wind for electric power production needs maintenance personal. So this option will not be a reliable power supply in remote areas for developing countries. Biomass energy sources are limited to small applications of individuals as an energy source but not

to produce electricity.

2. LIBYAN CONVENTIONAL ENERGY RESOURCES

The present energy supply in most countries cannot be considered as a sustainable sources of energy, as the energy costs are exploding, the sources are limited, and because of the environmental issues. For Libya the conventional sources of energy are limited to two sources.

1. Oil: with a total discovered resources estimated to 40 billion bbl.
2. Natural Gas: with a total discovered resources estimated to 1300 billion m^3.

The oil resources for Libya will not last more than 50 years as of today production and discovered resources, while the natural gas is expected to last more than that. Libya is an oil exporting country and most of the produced energy is exported. Table I shows the energy production by the year 2005.

TABLE I
Energy production by the year 2005

Type	Production	Consumption	Export
Natural gas	12 b m^3/y	3 b m^3/y	9 b m^3/y
Oil	0.6 b bbl/y	0.1 b bbl/y	0.5 b bbl/y
Electricity		20 T Wh/y	

In 2050 the price of oil barrel may reach more than 200 $. It is expected that Libya needs about 70 million barrels of oil per year for its electricity requirement and this will cost about 14 billion dollars per year. The question then is whether it will be better to sell the oil or to burn it. The answer will depend on the availability of other sources which can replace the conventional sources of energy like solar energy. The solar radiation in Libya is equivalent to a layer of 25 cm of crude oil per year on the land surface.

3. ENERGY REQUIREMENTS 2050

Libyan energy requirement scenario for 2050 is shown in Table II.

TABLE II
Scenario evolution for Libya

Type	2005	2050
Population	6 million	10 million
Electric power generation	3500 kWh/cap/y	5000 kWh/cap/y
Electricity demand	20 TWh/y	50 TWh/y
Water	6 billion m^3/y	10 billion m^3/y

4. RENEWABLE RESOURCES

Libya is located in the middle of North Africa with 88% of its area considered to be desert, the south is located in the Sahara desert where there is a high potentiality of solar energy which can be used to generate electricity by both solar energy conversions, photovoltaic, and thermal. The renewable energy sources for Libya according to the MED-CSP (Trans Mediterranean interconnection for concentrating solar power) scenario is shown in Table III, while the electric consumption and its sources in year 2050 is shown in Figure 1.

TABLE III
Renewable energy sources for Libya

Type	Potentiality
Solar electricity	140,000 TWh/y
Wind electricity	15 TWh/y
Biomass	2 TWh/y
Total	157,000 TWh/y

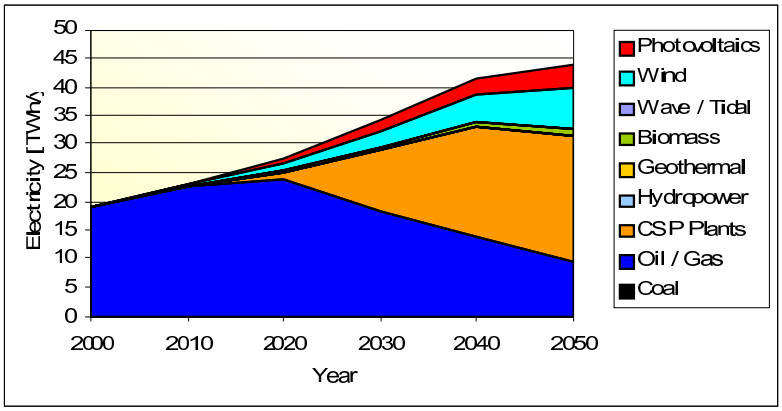


Fig. 1. Electricity consumption in Libya and supply resources.

5. RENEWABLE ENERGY APPLICATIONS

The use of renewable energies has been introduced in wide applications due to its convenience and being economically effective in many applications. The renewable energy sources used in Libya consists of photovoltaic conversion of solar energy, solar thermal applications, wind energy, and Biomass.

5.1. Photovoltaic

The use of PV systems started in 1976, and since then many projects have been developed for different sizes and applications. The first project put into work is a

PV system to supply a cathodic protection for the oil pipe line connecting Dahra oil field with Sedra Port. Projects in the field of communication were started 1980, when a PV system was used to supply energy to the microwave repeater station near Zella. Projects in the field of water pumping were started in 1983, when a PV pumping system was used to pump water for irrigation at El-Agailat. The use of PV systems for rural electrification and illumination started in 2003. PV systems found an increasing number of application types and their role grew considerably.

5.1.1. *PV in Microwave Communication Networks*

The Libyan Microwave communication network consists of more than 500 repeater stations. Only 9 remote stations were running by photovoltaic systems till the end of 1997 with a total peak power of 10.5 KWp. Remote stations in the eighties were running by diesel generators alone, while stations near general electric grid were powered by grid as the main supply and diesel generators as a backup. Nine stations were powered by photovoltaic systems at the beginning of 1980; four of these stations are still running after 26 years of work, the batteries which are of open type were replaced three times with an average lifetime of eight years. It was the technical and economical success of the PV systems that pushed the changing of all possible diesel stations to PV stations in the Libyan communication networks. The total number of stations running by PV in the field of communications currently exceeds 80 stations. The total installed photovoltaic peak power installed by the end of the year 2005 is around 420 KWp. Figure 2, shows the accumulated installed photovoltaic systems in the communication networks in the period 1980-2005 (Saleh Ibrahim *et al.*, 2003).

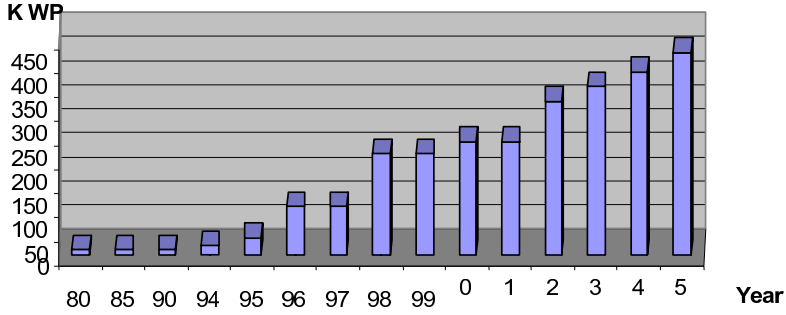


Fig. 2. The accumulated installed PV peak power 1980-2005

5.1.2. *PV in Cathodic protection*

In a previous study (Saleh Ibrahim *et al.*, 2004), it was shown that the cost of one KWh to supply a daily load of 15 KWh for a cathodic protection (CP) station is 1.4$ for a load which is located 5 Km from the 11 KV electric grid. In another study for the supply of a daily CP load of 7.5 KWh it was found that a PV system will be the most convenient solution at a distance of more than 1.2 Km from the 11 KV electric grid. Figure 3 shows the cost comparison to supply a CP station (Hibal *et al.*, 2004). The CP stations are usually far away from electric grid, a conclusion

out of this indicate that it is not feasible to use this type of source for this type of applications when a CP (15 KWh/day) station is located more than 2 Km from 11 KV transmission line.

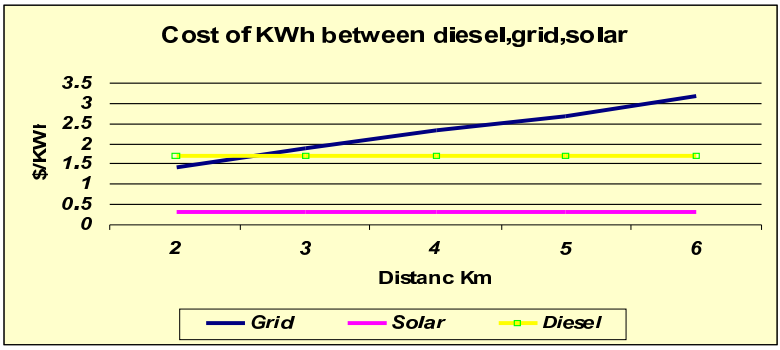


Fig. 3. Cost of KWh versus distance

From the cost comparison, it can be shown that the PV systems are the most economical choice; the choice will be more acceptable when considering real operations which include source failure. The first system in this field was put into work in 1976, the accumulated total power of PV systems in this field is second to PV systems in communications. The total number of PV systems in this field is around 300 by the end of 2005, with a total installed power of 540 KWp. PV technology is considered to be relatively new in developing countries; the problems we are facing have not to do with the technology but with people dealing with it. We are experiencing some vandalism issues: modules are often broken either by direct throwing stones on the module surfaces, or when shooting birds staying on the PV array.

5.1.3. *PV in Rural Electrification*

In most of the world the problem facing the electrification of low population areas or of regions far away from the existing electric networks is a financial one. It is so expensive to extend high voltage lines through desert to electrify few hundred inhabitants. In many low population countries electricity is only available in the cities and no electric network is used to power their rural areas. The Libyan national plans to electrify rural areas consist of electrifying scattered houses, villages, and water pumping. The PV supply systems for ten villages were introduced as a project to electrify remote areas (Saleh Ibrahim *et al.*, 2006). Some of these villages are

1. Mrair Gabis village as an example of scattered houses.
2. Swaihat village as an example of scattered houses
3. Intlat village as an example of scattered houses
4. Beer al-Merhan village as an example of scattered houses
5. Wadi Marsit village as an example of a village having diesel generator
6. Intlat village

The installation of photovoltaic systems started in the middle of 2003. The total number of systems installed by General Electric company of Libya (GECOL) is 340 with a total capacity of 220 kWp, while the ones installed by the Center of Solar

Energy Studies (CSES) and the Saharan Center are 150 with a total power of 125 KWp. The applications are: 380 systems for isolated houses, 30 systems for police stations, and 100 systems for street illumination. The total peak power is 345 KWp.

5.1.4. *PV for Water pumping*
Water pumping was considered as one of the best PV applications in Libya as remote wells which are used to supply water for human and live stock that are located in rural places. The water pumping projects consist of installing 35 PV systems with a total estimated peak power of 110 KWp.

Table IV shows the total installed PV capacity in Libya by the year 2005.

TABLE IV
Total installed PV capacity in Libya by the year 2005

Applications	Number of systems	Total power [KWp]
Communication	100	420
Cathodic protection	320	650
Rural Electrification	510	345
Water pumping	40	110
Total		1525

Gained experiences

From the collected data concerning the PV system performance in the local environment, it was proved that PV systems are highly reliable and cost effective. Some remarks drawn from past experience are (Saleh Ibrahim *et al.*, 2006):

1. No spare parts have been used for PV systems which were installed 26 years ago.
2. No failure has been registered for the systems installed 26 years ago.
3. Most of the PV systems had very low cost or no running cost.
4. Batteries have been changed after about ten years from installation.
5. People in developing countries should be informed about PV systems and their technology.
6. The average production energy for systems of 1.2 KWP is 6 KWh/day.
7. The AC option of electricity for rural electrification was the most convenient choice.
8. The closed type batteries option was the most convenient choice.

Social Impacts of PV Systems

Since photovoltaic technology is considered as relatively new; we are experiencing a lot of social changes for instance in the settlement of Bedouins. We are expecting that some small industries will be started. The availability of power supply will give a good chance to involve the populations of such remote areas increasing their knowledge and becoming familiar with the daily life of modern society. The availability of electrical power motivates the population to use more appliances like TV

sets, refrigerators that are normally in use in grid connected areas. As a result we have noticed a load increase in some houses which exceeded the maximum capacity of the PV supply systems. We also noticed that some population started to move back to these remote areas causing additional power loads. The reception of TV programs may change the habits of family live resulting in lower productivity.

We have not experienced any vandalism. The only problem reported in one of the system in which the inverter stopped because of overload may be considered to be due to the equipment itself.

5.2. Thermal Conversion

The use of solar heaters started in 1983 with a pilot project which included 10 systems. Since then about 2000 additional solar heaters have been installed in Libya.

Water heating energy consumption is about 12% of the national electricity production but the use of solar heaters has not been spread in all the country due to

1. No national or personal industry has been established for local individuals.
2. Lack of information for the people.
3. Low electric energy tariff.

5.3. Wind Energy in Libya

Wind energy was utilized for water pumping in many oasis since 1940. The use of this energy has not been developed extensively because the wind-mills need some maintenance from time to time.

In 2004 measurements of the wind speed statistics has been conducted and showed that there is a high potentiality for wind energy in Libya. The average wind speed at a 40 meter height is between 6- 7.5 m/s. Figure 4 shows the average wind speed measured in different locations of the Libyan coast area.

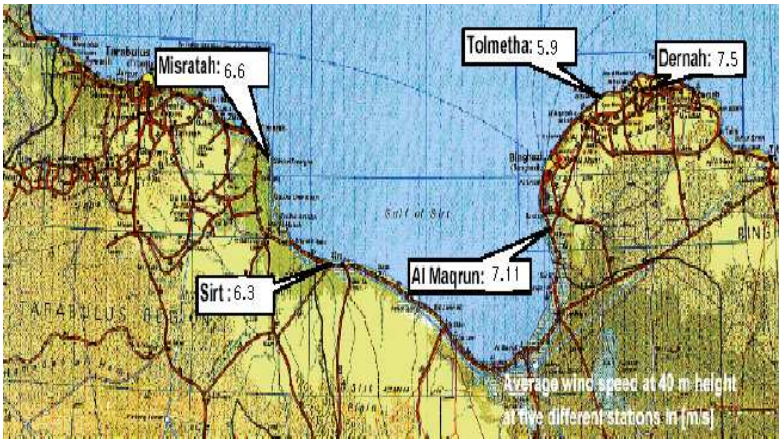


Fig. 4. Average wind speeds in m/s measured in 5 locations along the Libyan coast.

The use of wind energy for electricity production has not been started yet in

Libya, but a project was contracted for installing 25 MW as a pilot project to be erected in two years time. A project to prepare two Atlas that provide fast access to reliable solar and wind data through out Libya is also been contracted for. The Atlas will allow for accurate analysis of the available wind and solar resources anywhere in Libya, and will be therefore very important for planning profitable wind farms and solar projects.

6. FUTURE PROSPECTS

Based on the past gained experience, a national Renewable Energy (RE) plan has been proposed with the aim of bringing RE into the main stream of the national energy supply system with a target contribution of 10% of the electricity demand by the year 2020. The proposed plan call for a wide spectrum of Renewable energy applications. Table V shows the planed renewable energy contribution in Libya from different sources.

TABLE V
Planed renewable energy production in Libya in 2020.

Total	Technology
10 MWp	PV
150 MW	Wind
20,000 m^3	Thermal Water heating
20 MW	Thermal electricity
20,000 m^3	Thermal Desalination
20 KW	Hydrogen

Due to high potential of renewable energy sources in Libya and its location near the energy market in Europe, it is possible to plan in the future to generate electricity from renewable sources in Libyan southern regions and to deliver it to Europe. Figure 5 shows a map for future electricity supply scenario from renewable sources.

7. CONCLUSION

Libya can be considered as a place with high potentiality for renewable energy production. The use of a stand-alone PV power supply in communications, cathodic protection, rural electrification, and water pumping, is justified. Finally we can say that:

1. There is a good potentiality for PV systems which can be used in different applications.
2. Photovoltaic systems for supplying electrical energy to remote areas are justified based on economical and technical reasons.
3. Social changes have been noticed in the villages which have been electrified.
4. There is a potentiality of renewable (solar & wind) energy which can be used in different applications.

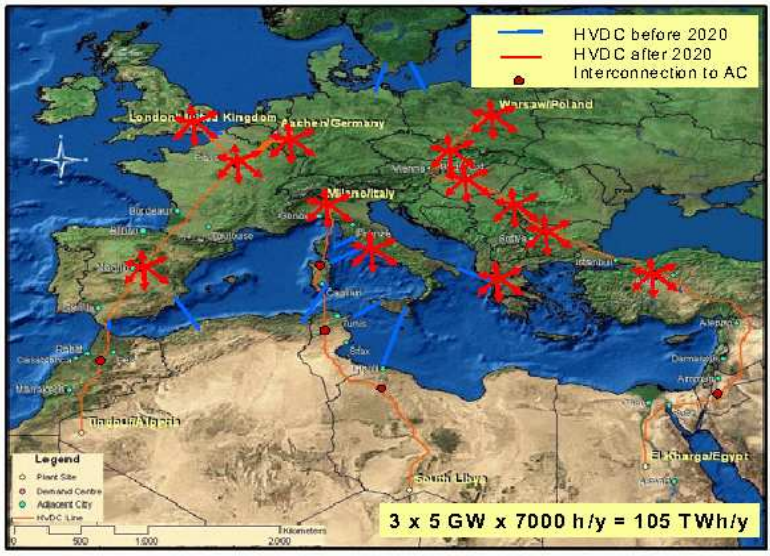


Fig. 5. Map for future electricity supply scenario from renewable sources MED-CSP.

5. A national plan has been adopted to rise the share of renewable energy to 10% by the year 2020.
6. Number of projects are progressing in the field of renewable energy

References

Hibal, *et al.*: 2004, M.Sc. Thesis, Faculty of Engineering, Al-Fateh University, Libya

Saleh Ibrahim, I. M.: 1993, in The forth Arab International Solar Energy Conference, Amman, Jordan

Saleh Ibrahim, I. M., *et al.*: 1998, Renewable Energy Journal, **14**, 135

Saleh Ibrahim, I. M., *et al.*: 2003, Photovoltaic conversion in Telecomunication Network in Libya, in Renewable energy and Environment Conference, Brak, Libya (in Arabic)

Saleh Ibrahim, I. M., *et al.*: 2004, Experiences with photovoltaic systems used for cathodic protection in Arabian gulf oil company, in The twentieth conference and exhibition on photovoltaic solar energy conversion, Paris, France

Saleh Ibrahim, I. M., *et al.*: 2006, Photovoltaic in libya applications, and evaluations, in International Conference on renewable energy for developing countries, Washington


Well Come t
2006 . 3

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

COMPUTER SIMULATION FOR CURRENT DENSITY IN PN-SILICON SOLAR CELLS

FATHI Y. R. EL FAITURI and ALI Y. DARKWI
University of Garyounis, Benghazi-Libya
E-mail: Fathi_ff2002@yahoo.com
E-mail: alidarkwi@yahoo.com

Abstract. Numerical analysis of electric field and potential profile in depletion layer of an abrupt Si p-n junction solar cells has been made by solving Poisson's equation. The variation of depletion layer as well as the electrostatic potential and electric field with forward bias voltage were also analyzed. The transport and continuity equations for the charge carrier were solved numerically to obtain the current density. The evaluated current density considered both the diffusion in neutral regions of (p and n) and drift within depletion layer. The influence of lifetime and surface recombination speed of charge carries in the efficiency of the cell were also discussed.

Key words: pn junction, electron and hole life time, surface recombination speed

1. Introduction

The conversion of sunlight directly to electric power has been dominated by solid state junction devices. The junction between different material can set up an internal electric field, which is responsible for separation of charge pairs (electron and hole) that are generated by incident photons. The distribution of electric field in the depletion layer can be obtained by solution of Poisson's equation. Although this equation does not appear to have an analytical solution, many articles (Green, 1982; Pierret, 1996; Zambuto, 1989) deal with an analytical treatment, which allows an easy interpretation of physical results and direct control on the involved parameters. A limitation of analytical models lays in the need of simplified assumption on the physics of the device, which may only be accepted in order to reduce the complexity in the calculation. Numerical treatment offers a deeper comprehension of the structure, achieving a complete control on the various parameters and defining their role in the device operation. A very important contribution in the numerical analysis of electrostatic potential and electric field in a depletion layer of linearly-graded p-n junction has been introduced by Kennedy and O'Brien (1967). In their analysis, the space-charge density modifications arise from impurities as well as from free charge of electrons and holes. In particular, our analysis is based on the numerical solution of Poisson's equation in the depletion layer of an abrupt Si p-n junction. The space-charge density ρ involved in the calculation considers both the charge due to impurities and free charge of electrons and holes, however, some articles consider only the charge due to impurities (Green, 1982; Tsaur *et al.*, 1972). The calculation has been done for equilibrium (V bias =0) and under forward bias voltage. The aim of this calculation is focused on an accurate determination of space charge density and the behavior of depletion layer under

bias voltage. A numerical model has been developed to find the current density. Based on physical consideration on minority carriers, the transport and continuity equations were solved numerically. Flexibility of the model lies in the capability of obtaining the carrier distribution and hence, the current density varying with material parameters or structure parameters. In this analysis we have tried to show how the efficiency of the cell can be enhanced by changing the parameters involved in the calculation.

2. Electric Field Within Depletion Layer of An Abrupt p-n Junction

The distribution of electric field within the depletion layer can be analyzed by solving Poisson's equation

$$\frac{d^2\phi}{dx^2} = -\frac{1}{\epsilon}\rho(x,\phi) \tag{1}$$

Where ϕ is the electric potential, ϵ is dielectric constant, ρ is the space charge density, and x is the depth in the depletion layer. The space charge density in the depletion layer is given by

$$\rho = q\left[N_d(x) - N_a(x) + p(\phi) - n(\phi)\right] \tag{2}$$

where $p(\phi)$ and $n(\phi)$ are the concentration of free charge of holes and electrons respectively, $N_d(x)$ and $N_a(x)$ are the concentration of ionized donors and acceptors(Sze, 1981).

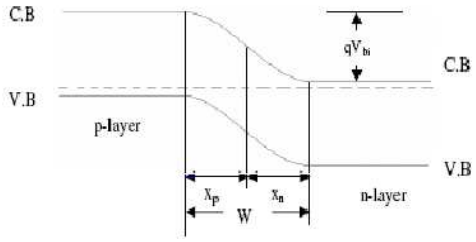


Fig. 1. The pn junction at equilibrium.

Within the depletion layer, Eq.(1) was solved numerically. The left side of the equation can be approximated by

$$\frac{\phi_{i-1} - 2\phi_i + \phi_{i+1}}{h^2} - err(h) = -\frac{\rho(x_i,\phi_i)}{\epsilon} \tag{3}$$

Where h is an increment ($h = x_{i+1} - x_i$) and the part $err(h)$ is the local error. When h is very small, this term can be neglected. Hence, eq.(3) becomes

$$\phi_{i-1} - 2\phi_i + \phi_{i+1} + \frac{h^2\rho(x,\phi)}{\epsilon} = 0 \tag{4}$$

Where $i = 1, \ldots, N$.

With a proper physical boundary conditions (EL Faituri, 2004), the above equation was solved by iteration method. The space-charge density, electrostatic potential and electric field are shown in the following figures. For an abrupt p-n junction, the doping profile of the impurity is assumed to be uniform throughout each semiconductor (p and n) up to the junction. Within depletion layer, the doping profile is shown in Fig. 2 curve (a). While the free negative $n(\phi)$ and positive $p(\phi)$ charge density within depletion layer which tails off as one proceed toward the junction is illustrated in curve(b) and (c) respectively. Since the density of free charges $n(\phi)$ and $p(\phi)$ were low comparing to the density of donor N_d and acceptor N_a, the shape of net charge density is overruled by impurities (N_a and N_d) as shown in curve (d); however, the contribution of free charges become clear in reducing the space-charge density. The corresponding electrostatic potential and electric field is shown in curve (e) and (f) respectively. The analysis was carried out under equilibrium conduction (bias voltage V=0). The continuity of electrostatic potential and electric field at $x = 0$ justifies the condition of charge neutrality (Zambuto, 1989), for which $N_a x_p = N_d x_n$. Within the depletion layer, the electric field is negative, has maximum value at x=0 and exhibits a linear variation with position. The negative profile of the electric field simply originates from the choice of the boundary condition in one dimension for which the electric field is pointed in the negative x-direction, while the linear variation of electric field with position is due to charge distribution. Since there is no charge outside the depletion layer, the electric field becomes zero in the neutral region of n and p.

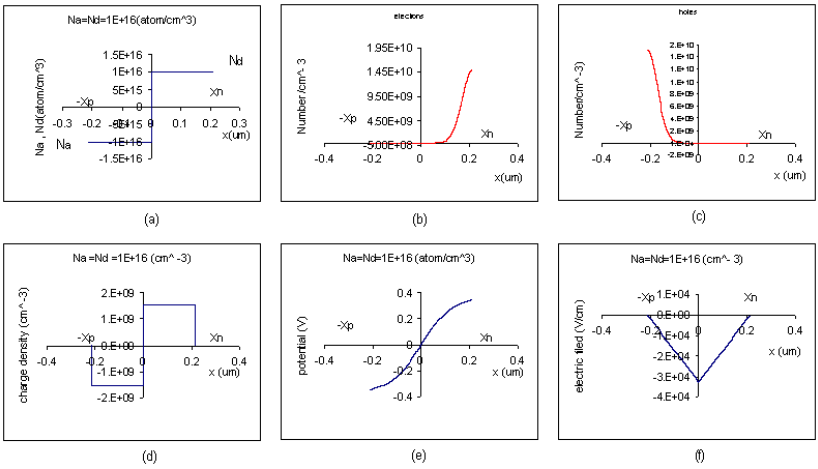


Fig. 2. Curves for an abrupt p-n junction in silicon at 300 K, Curve (a) shows the density of donor ($N_d = 10^{16} cm^{-3}$) and acceptor ($Na = 10^{16} cm^{-3}$) atoms within the depletion layer as a function of x, curves (b) and (c) show the density of free negative charge ($n(x)$) and positive ($p(x)$) within the depletion layer; Curve (d) shows the net charge density ρ(x) within the depletion layer related to curves (a),(b)and (c); Curve (e) shows the electric potential within the depletion layer as a function of x; curve (f) shows the electric field within the depletion layer as a function of x.

As a matter of fact, the solar cell operates in forward bias voltage, therefore we are particularly interested in how the electrostatic variables (potential and field)

could change under this biasing. In forward bias case, the barrier between p-and n-side is reduced to the relation $V_{bi} - V$ where $V > 0$. The lower barrier permits a large portion of majority carriers to cross the junction into the adjacent semiconductor, where they now become minority. The pileup of minority carriers in the immediate vicinity of the depletion layer results in concentration gradient, which causes diffusion of the carriers away from the junction toward a quasineutral region, where they recombine with the majority. Hence, the diffusion causes a reduction of charge along the two sides of the junction (Zambuto, 1989). Therefore, the depletion layer becomes narrow, the electrostatic potential and electric field decrease everywhere inside the depletion layer. Fig.3. shows, the electrostatic potential and electric field profiles within the depletion layer at forward bias voltage 0.2 and 0.4 volt.

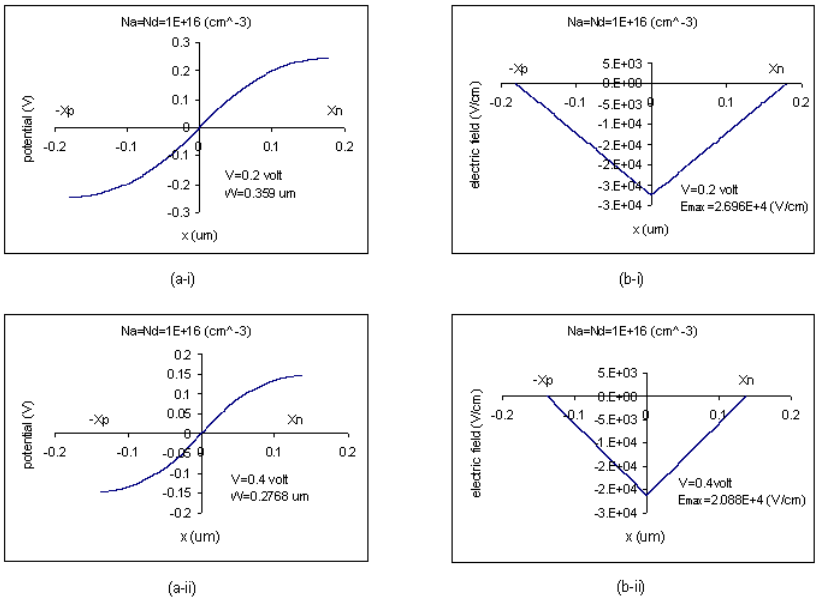


Fig. 3. Curves for an abrupt p-n junction in silicon at 300 K, under forward bias voltage. Density of donor and acceptor atoms is $N_d = 10^{16}$ and $N_a = 10^{16}$ respectively: (a) the electric potential within the depletion layer as a function of x, (b) the electric field within the depletion layer as a function of x . Where: W, is the width of the depletion layer, $E_{\max}$, is the maximum value of the electric field and V is the forward bias voltage.

3. Analysis of Current Density

The current density was obtained by the solution of the electron and hole continuity equations

$$-\frac{1}{q}\frac{dJ_n(x)}{dx} = G(x) - R(x) \tag{5}$$

$$\frac{1}{q}\frac{dJ_p(x)}{dx} = G(x) - R(x) \tag{6}$$

where

$$J_n = q\mu_n n(x)E(x) + qD_n\frac{dn(x)}{dx} \tag{7}$$

and

$$J_p = q\mu_p p(x)E(x) - qD_p\frac{dn(x)}{dx} \tag{8}$$

are the electron and hole conduction current densities. $G(x)$ is the generation rate, $R(x)$ is the recombination rate, μ_n and μ_p are the mobility of free electrons and holes, D_n and D_p are the diffusion coefficients of electrons and holes respectively.

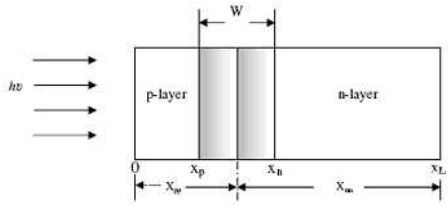


Fig. 4. p-n junction solar cell illuminated from p-side.

The generation rate $G(x)$ of charge carriers in the semiconductor layer at the depth x is given by;

$$G(x) = \int_0^{\lambda_{max}} [1 - R(\lambda)]\alpha(\lambda)N_{ph}(\lambda)e^{-\alpha(\lambda)x}d\lambda \tag{9}$$

where $\alpha(\lambda)$ is the absorption coefficient, $N_{ph}(\lambda)$ is the incident photon flux, $R(\lambda)$ is the reflection coefficient at front surface. λ_{max} is the maximum wavelength determined by the condition that, each photon with energy $h\nu \geq E_{gap}$ could generate electron-hole pairs. The recombination rate R(x) is described by the lifetime of electrons (τ_n) and holes (τ_p) as follows (Fonash, 1981);

$$R(x)_n = \frac{\Delta n(x)}{\tau_n} \tag{10}$$

$$R(x)_p = \frac{\Delta p(x)}{\tau_p} \tag{11}$$

where $\Delta n(x)$ and $\Delta p(x)$ are the excess electron and hole concentrations respectively. The electric field within the neutral region of the p-and n layer is zero, hence, using equations (7) and (8), the equations (5) and (6) can be reduced into second-order differential equations of the form

$$D_n\frac{d^2\Delta n(x)}{dx^2} - \frac{\Delta n(x)}{\tau_n} + G(x) = 0 \tag{12}$$

$$D_p \frac{d^2 \Delta p(x)}{dx^2} - \frac{\Delta p(x)}{\tau_p} + G(x) = 0 \tag{13}$$

Eqs.(12) and (13) are iteratively solved with appropriate boundary conditions (Sze, 1981; Donald, 1992). The current densities due to diffusion of charge carriers are analyzed using (Sze, 1981).

$$J_n = qD_n \frac{d\Delta n}{dx}|_{x=x_p} \tag{14}$$

and

$$J_p = -qD_p \frac{d\Delta p}{dx}|_{x=x_n} \tag{15}$$

Within the depletion layer, the electric field is generally high. We assume that, the photogenerated carriers are accelerated out of the depletion layer before they can recombine. Thus, the current density due to depletion layer is

$$J_{dep} = q \int_0^{\lambda_{max}} [1 - R(\lambda)] N_{ph}(\lambda) e^{-\alpha(\lambda) x_p} [1 - e^{-\alpha(\lambda) W}] d\lambda \tag{16}$$

The total photocurrent density is then

$$J_{ph} = J_n + J_p + J_{dep} \tag{17}$$

and the total current density

$$J(V) = J_{ph} - J_{dark} \tag{18}$$

where $J_{dark} = J_0[e^{\frac{qV}{kT}} - 1]$. J_0 is the saturation current density.

4. Interpretation of Numerical Results

Figure 5 presents the numerical result on current density profiles as a function of charge carriers lifetime. Curves in panel (a) show the current density calculated for two solar cells with different hole lifetime while electron lifetime is kept at $\tau_n = 10^{-6}$ sec. The short circuit current density (J_{sc}) and efficiency (E_{ffi}) are reduced as the lifetime decreases. In fact, when the charge carrier lifetime decreases, the diffusion length ($L = \sqrt{D\tau}$) becomes small and hence a few carriers can contribute in the diffusion process. This causes reduction in short circuit current density. The curves in panel (b) was obtained by keeping the hole lifetime fixed at $\tau_p = 10^{-6}$sec while changing the electron lifetime. We can see that both (J_{sc}) and (E_{ffi}) present an appreciable reduction which imply the same conclusion pointed above.

Figure 6 shows the current density characteristics for a solar cell as a function of surface recombination speed of charge carrier. Curves in panel (a) and (b) show the current density when varying then electron surface recombination speed (S_n) keeping the hole surface recombination speed (S_p) unchanged. As the electron surface recombination speed decreases, the short circuit current (J_{sc}) and the efficiency (E_{ffi}) increase in agreement with the result obtained by (Yasutake *et al.*, 1994).

TABLE I
parameters used for numerical calculations

Parameter	Value
T	Temperature, 300K
ϵ_s	Dielectric constant for silicon, 1.0610^{-10} Fm^{-1}
D_n	Diffusion coefficient of electrons, D_n=34.91(cm^2/s)
D_p	Diffusion coefficient of holes, D_p=12.41(cm^2/s)
X_{pp}	Width of p-layer at zero bias voltage, 5 μ m.
W	Width of depletion layer at zero bias voltage, 0.0201μ m
X_{nn}	Width of n-layer at zero bias voltage, 245μ m
$\alpha(\lambda)$	Absorption coefficient of Si (Aspnes *et al.*, 1983)
$R(\lambda)$	Reflection coefficient at the front surface (EL Faituri, 2004)
$N_{ph}(\lambda)$	Incident Photon flux for AM1, (100 $\mathrm{mW/cm}^2$) (Chopra and Das, 1983)
τ_n, τ_p	Lifetime of electrons and holes respectively, (from 10^{-6} to 5×10^{-4} sec) (Duffie and Beckman, 1974)
S_n, S_p	Surface recombination speed for electrons and holes respectively (10^2, 10^5 cm/sec) (Datta *et al.*, 1994)

Curves in panel (c) and (d) are obtained by keeping the electron surface recombination speed constant and changing the hole surface recombination speed. For the case of long base layer (n-layer), there is no big change in (J_{sc}) and (E_{ffi}). This because, the rate of generation and collection of charge carriers remain the same at the back of the base layer. However, when the base layer becomes short, the distribution of generated charge carriers becomes high (EL Faituri, 2004). For short base layers the (J_{sc}) and (E_{ffi}) become high as the surface recombination speed decreases. The result corresponds to what indicated in some articles (Datta *et al.*, 1994; Yasutake *et al.*, 1994), for which the efficiency of thin silicon solar cell increases while (S_n) and (S_p) become small.

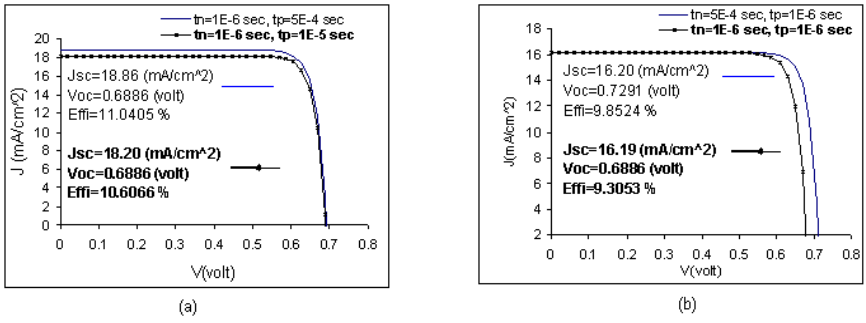


Fig. 5. Calculated J-V characteristics as a parameter of lifetime. Curves in panel (a) consider cells with the same electron lifetime but different hole lifetime, curves (b) represent the inverse situation, same τ_p but different τ_n.

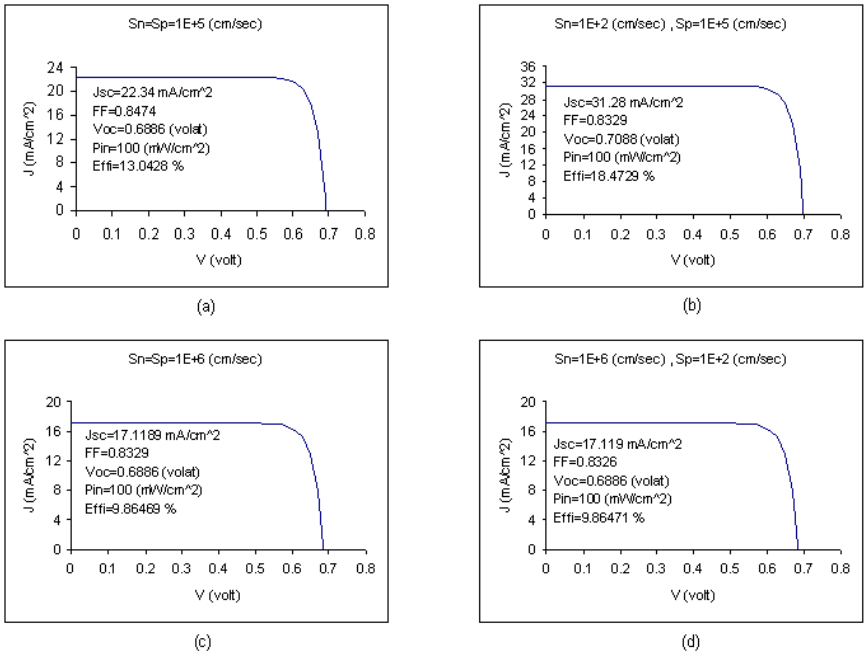


Fig. 6. Calculated J-V characteristics as a parameter of surface recombination speed. Curves in panels (a) and (b) consider cells with the same hole surface recombination speed but different electron surface recombination speed, curves (c) and (d) are obtained for the inverse situation, same S_n but different S_p.

5. Conclusion

In this paper, an accurate analysis of charge density in the depletion layer has been considered. It has been shown that, the density of charge within the depletion layer decreases when free charges (electrons and holes) are involved in the calculations. Because of the low density of free charges compared to the charge density due to impurities, the shape of space-charge density is overruled by impurities. From J-V characteristics, the effect of surface recombination speed of charge carriers, shows that, the efficiency increases as the surface recombination speed decreases. Therefore, for good solar cell fabrication, one can chose materials that have low surface recombination speed in the front and back contact of the cell.

References

Aspnes, D. E. and Studna, A. A.: 1983, *Physical Review B*, **27**, 985
Chopra, K. L. and Das, S. R.: 1983, *Thin Film Solar Calls, Plenum press*, New York, P. 508
Datta, S. K., Mukhopadhyay, K., Pal, P. K., Saha, H.: 1994, *Solar Energy Materials and Solar Cells*, **33**, 483
Duffie, J. A. and Beckman, W. A.: 1974, *Solar Energy Processes*, John Wiley and Sons, Inc.
EL Faituri, F. Y.: 2004, Msc. Thesis, University of Garyounis.
Fonash, S. J. : 1981, *Solar Cell Device Physics*, Academic press, Inc, P. 150
Green, M. A.: 1982, *Solar Cells*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, P. 66

Kennedy, D. P., O'Brien, R. R.: 1967, *IBM Journal of Research and Development* **11**, no.3, 252
Donald A. N.: 1992, *Semiconductor Physics and Devices*, IRWIN, Inc, P. 247
Pierret, R. F.: 1996, *Semiconductor Device Fundamentals*, Addison-Weley Publishing Company, Inc.
Sze, S. M. : 1981, *Physics of Semiconductor Devices*, by John Wiley and Sons, Inc.
Tsaur, S. C., Milnes, A. G., Sahai, R. and Feucht, D. L.: 1972, in *Symposium on GaAs. Inst. of Phys. and Phys. Soc., London.*, 156
Yasutake, K., Chen, Z., Pang, S. K., and Rohatgi, A.: 1994, *J. Appl. Phys.*, **75**, (4), 2048
Zambuto, M.: 1989, *Semiconductor Devices*, Mc Graw-Hill Book Company, P. 183


Winzrik
Group

International Symposium on Solar Physics and Solar Eclipses (SPSE) 2006
R. Ramelli, O. M. Shalabiea, I. Saleh, and J.O. Stenflo, eds.

CONCLUSIONS OF THE SYMPOSIUM

The International Symposium on Solar Physics and Solar Eclipses (SPSE 2006) was a great success. A large number of prominent scientists from all over the world came with complex equipment to carry out advanced scientific experiments and take part in a high-quality conference on both solar physics and solar energy utilization. The observing conditions and infrastructure support at the eclipse camp near Waw an Namos were excellent. Here is a brief summary of some of the main results:

- 94 scientists from 11 countries took part in the symposium with a series of high-level talks on topics from history of astronomy, eclipse observations, coronal and space physics, fundamental physics, instrumentation, and solar energy utilization. This attendance on such a remote site in the midst of the Sahara desert exceeded the original expectations.
- The symposium brought together, for the first time, scientists representing the two disciplines of solar physics and solar energy utilization.
- The conference concluded that Libya, with its sunny climate, large land space and low population density is in a unique position in the world to develop the almost untapped source of solar energy. In view of the end of the oil era within the coming decades a major shift from fossil fuels to sustainable energy is urgently needed, and solar energy is the key ingredient in a sustainable energy future.
- A number of advanced scientific experiments from different countries and with different scientific aims were carried out with great success, giving unique new insights about the Sun.

AUTHOR INDEX


КБ-2
Pilots Association