Observations of the High-Energy Peaked BL Lac object H 1426+428 with the Solar Tower Atmospheric Cherenkov Effect Experiment

Carsten Mueller

Department of Physics

McGill University, Montréal

April 2008

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

In memory of STACEE
. . . and all the sleepless nights she caused

The mathematical sciences particularly exhibit order, symmetry, and limitation; and these are the greatest forms of the beautiful.

– Aristotle, *ca.* 330 B.C. (*Metaphysica 3-1078b*)

ABSTRACT

The Solar Tower Atmospheric Cherenkov Effect Experiment (STACEE) was an atmospheric Cherenkov telescope operational until June 2007, which detected cosmic gamma rays by means of the wavefront-sampling technique. Using 64 of the large heliostats available at the National Solar Thermal Test Facility (NSTTF) near Albuquerque, New Mexico, its total mirror area of $2378\,\mathrm{m}^2$ allowed it to achieve energy thresholds between 150 and 200 GeV. Following a review of the field of gamma-ray astrophysics and Active Galactic Nuclei, this work provides a detailed description of the detector, along with an introduction to simulation and data analysis techniques. During the springs of 2003 and 2004, STACEE observed the High-Energy-Peaked BL Lac object H 1426+428, an established emitter of gamma rays in the TeV energy range. A full analysis of these data reveals that STACEE did not detect a statistically significant gamma-ray excess from H 1426+428 in either data set. With the help of detailed detector simulations, upper limits on the integrated gamma-ray fluxes, Φ_{int}, at the 95% confidence level are obtained for both observing seasons: $\Phi_{\mathrm{int}}(E > 163\,\mathrm{GeV}) < 1.83 \cdot 10^{-6}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$ and $\Phi_{\mathrm{int}}(E > 165\,\mathrm{GeV}) < 0.78 \cdot 10^{-6}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$ for 2003 and 2004, respectively. The derived upper limits are discussed in the context of the results obtained by other gamma-ray observatories.

ABRÉGÉ

Le télescope à effet Čerenkov atmosphérique STACEE (Solar Tower Atmospheric Cherenkov Effect Experiment), en opération jusqu'en juin 2007, servait à détecter des rayons gamma d'origine cosmique en utilisant la technique d'échantillonnage du front d'onde. L'utilisation de 64 héliostats de la centrale solaire NSTTF (National Solar Thermal Test Facility) située près d'Albuquerque au Nouveau Mexique permettait au télescope d'atteindre des seuils d'énergie entre 150 et 200 GeV. Cette thèse commence par une revue dans le domaine de l'astrophysique des rayons gamma et des noyaux actifs galactiques, puis présente une description détaillée du détecteur STACEE, de même qu'une introduction aux techniques de simulation et d'analyse de données. Entre mars et juin 2003 ainsi qu'entre mars et juillet 2004, STACEE a observé l'objet HBL H 1426+428, un émetteur avéré de rayons gamma TeV. L'analyse complète des données recueillies révèle que STACEE n'a permis de détecter un excès de rayons gamma statistiquement significatif provenant de H 1426+428 dans aucun des deux ensembles de données. Des simulations détaillées de l'effet du détecteur ont permis d'obtenir des limites supérieures du flux intégré de rayons gamma, $\Phi_{\rm int}$, à un niveau de confiance de 95% pour les deux saisons d'observation: $\Phi_{\rm int}(E > 163\,{\rm GeV}) < 1.83 \cdot 10^{-6}\,{\rm m}^{-2}\,{\rm s}^{-1}$ pour les observations effectuées en 2003, et $\Phi_{\rm int}(E > 165\,{\rm GeV}) < 0.78 \cdot 10^{-6}\,{\rm m}^{-2}\,{\rm s}^{-1}$ pour celles de 2004. Les limites supérieures ainsi obtenues sont comparées aux mesures de flux enregistrées par d'autres détecteurs, et des implications possibles sont discutées.

ACKNOWLEDGMENTS

Writing this doctoral thesis was really just the conclusion of many years of learning and working in the exciting field of gamma-ray astrophysics. Being part of the STACEE collaboration was a very rewarding experience. As the group was relatively small, every member had ample opportunity to work with the detector, take data, analyse data, and participate in collaboration meetings. I have pleasant memories of the long data-taking shifts in New Mexico, which provided me with a lot the hands-on experience. I thank my supervisor Professor Ken Ragan for giving me the opportunity to do all this, and also for his guidance, ability to inspire enthusiasm and excitement about the field, and, not to be forgotten, patience.

There are many others without whom the work summarised by this thesis would not have been possible. David Hanna, like an unofficial supervisor, helped me with advice on data analysis and with his critical thinking. Jim Hinton provided me with software tools that helped me get started, and John Kildea was always available for talking about STACEE and proof reading parts of this thesis. I enjoyed working with all STACEE graduate students, post-docs and faculty, from McGill and from the other institutions in the U.S. and Canada. The success of a collaboration such as STACEE depends on contributions from each and every member, and thus I give credit to all of them for making the experiment work, keeping it running and providing specialised software. Among the former graduate students, Lowell Boone, Douglas Bramel, Pascal Fortin, Thomas Lindner and Richard Scalzo deserve special acknowledgement, because their contributions to the experiment greatly facilitated my task.

At McGill, I am greatly indebted to Paul Mercure for doing an excellent job as system administrator, and for allowing me to tap into the departmental processing power and use Terabytes worth of disk storage space for my simulations.

I thank the staff of the NSTTF at Sandia National Laboratories for always making us "astronomers" feel welcome at the site, taking care of missing badges (Cheryl Ghanbari and Loula Killian), promptly repairing defective heliostats and giving us rides to the tower (Mike Usher), or installing and repairing STACEE infrastructure (Kye Chisman and J.J. Kelton).

Last but not least, I am grateful to my parents for their continuing support throughout my long years of graduate studies.

STATEMENT OF ORIGINAL WORK

From the time I joined the STACEE collaboration (in 2000) on, I was involved in simulation projects. After writing a toy Monte-Carlo for elucidating the limits of STACEE's angular reconstruction capabilities, I took over responsibility for STACEE's electronics simulation (`Elec`). In the following years, I turned the electronics simulation from a crude framework which contained a few major bugs to a rather comprehensive simulation program capable of producing fake data for all possible STACEE trigger modes, which looked very much like real data. A few of the many additions to the program were initially written by collaborators, which I then integrated into the program.

From 2002 on, I was also the maintainer of STACEE's version of the CORSIKA air shower simulation code. I configured a new version of CORSIKA for STACEE use, adding support for a choice of different hadronic interaction models (including FLUKA) and different atmospheric models. I eventually modified the output routine to suit STACEE's increasing demand for accuracy, as well as the routines which allow other programs to read CORSIKA output. I adapted and expanded the simulation driver script, and wrote a higher level script that streamlined the running of simulations. Last but not least, I was in charge of keeping all simulation software up to date and functioning as a pipeline, and in this function was involved in the development of other simulation helper programs and writing a rather extensive amount of documentation. In the process of maintaining the simulation programs I also corrected several major errors.

I performed a large variety of simulation studies, some of which had the aim of finding the optimum configuration of the detector for gamma-ray observations, including optimisations of the various trigger parameters and of the canting parameter.

From the years 2001 to 2006, I accumulated a total of almost six months worth of observing, calibration and debugging shifts at the experiment site in New Mexico. The tasks performed there ranged from a multitude of normal observing procedures which included setting up the detector, running the DAQ, controlling the heliostats and the telescope-based atmospheric monitor system, to testing and calibrating detector components, realigning all optical components of STACEE, and locating and rectifying problems with the detector electronics.

For the data analysis, I customised parts of the standard STACEE analysis software, and generated all necessary Monte-Carlo results.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

ACT	Atmospheric Cherenkov Telescope
ADC	Analog-to-Digital Converter
AGN	Active Galactic Nucleus
ASL	Above Sea Level
BL Lac	BL LACertae (class of blazars without emission lines)
CAMAC	Computer Aided Measurement And Control (module standard)
CERN	European Laboratory for Particle Physics
CM	Centre-of-Mass
CMB(R)	Cosmic Microwave Background (Radiation)
DAQ	Data AcQuisition
DTIRC	Dielectric Total Internal Reflection Concentrator
EAS	Extended Air Shower
EBL	Extragalactic Background Light
EC	External Compton (leptonic model of γ-ray emission from blazars)
ESA	European Space Agency
eV	electron Volt (1 eV = $1.602177 \cdot 10^{-19}$ J)
FADC	Flash-ADC (continuously sampling analog-to-digital converter)
fan-in	Electronic module which outputs the sum of two or more input signals
fan-out	Electronic module which outputs two or more copies of the input signal
FOV	Field Of View
FSRQ	Flat-Spectrum Radio Quasar
FWHM	Full Width at Half-Maximum
GRB	Gamma-Ray Burst
GTC	Global Trigger Condition (STACEE Level-2 trigger multiplicity)
HBL	High-energy peaked BL Lacertae
HA	Hour Angle
HE	High Energy ($\sim 10^8$ - 10^{10} eV)
HV	High Voltage
IACT	Imaging Atmospheric Cherenkov Telescope
IGMF	InterGalactic Magnetic Field
IR	InfraRed range of electromagnetic radiation (700 nm $\lesssim \lambda \lesssim$ 1 mm)

LAB	LABoratory frame-of-reference
LBL	Low-energy peaked BL Lacertae
LED	Light-Emitting Diode *or* Leading-Edge Discriminator
LTC	Local Trigger Condition (STACEE Level-1 trigger multiplicity)
$M_\odot$	Solar mass ($1\ M_\odot = 1.98892 \cdot 10^{30}$ kg)
MADDOG	McGill digital delay and trigger system
MCS	Master Control System (heliostat motion control hardware)
MHD	Magneto-HydroDynamics
NASA	NAtional Space Agency
NIM	Nuclear Instrumentation Module (module standard)
NSB	Night Sky Background
OVV	Optically Violently Variable quasar (blazar-type FSRQ)
pc	Parsec ($1\ \mathrm{pc} = 3.085678 \cdot 10^{16}$ m $\simeq 3.262$ ly)
PE	PhotoElectron
PMT	PhotoMultiplier Tube, short *Phototube*
PSF	Point-Spread Function (measure of the resolution of a detector)
QED	Quantum ElectroDynamics
Quasar	QUAsi-StellAr Radio source (radio-loud AGN with no visible host galaxy)
QSO	Quasi-Stellar Object (radio-quiet AGN with no visible host galaxy)
RA	Right Ascension
RF	Radio Frequency
register	Electronic module for storage of numbers or codes
σ	Standard deviation (used as a unit)
SAM	STACEE Atmospheric Monitor
SED	Spectral Energy Distribution
SNR	SuperNova Remnant
SPB	Synchrotron Proton Blazar (hadronic model of γ-ray emission from blazars)
SSC	Synchrotron Self-Compton (leptonic model of γ-ray emission from blazars)
SSRQ	Steep-Spectrum Radio Quasar
TDC	Time-to-Digital Converter
TOF	Time-Of-Flight
UHE	Ultra-High Energy (above 10^{14} eV)
UV	UltraViolet range of electromagnetic radiation ($4\ \mathrm{nm} \lesssim \lambda \lesssim 380\ \mathrm{nm}$)
VHE	Very-High Energy (10^{10} - 10^{14} eV)
VME	Versa Module Europe (module standard)

CHAPTER 1

HIGH-ENERGY ASTROPHYSICS

The science referred to as *high-energy astrophysics* is a relatively recent branch of physics. It encompasses the physics devoted to the experimental study of the highest energy phenomena in the universe and their theoretical interpretation. On the experimental side, it is concerned with detecting particles from space which have been accelerated to very high energies, often many orders of magnitude higher than those that can be produced by the largest particle accelerators on Earth. It combines our knowledge about the smallest fundamental particles and the largest structures in the universe. While the current and future findings of high-energy astrophysics are of paramount importance for our understanding of conventional physics under extreme conditions, of cosmology, and of still unknown, exotic physical processes beyond the Standard Model of particle physics, the technical challenges inherent to the detection of highly energetic particles from celestial objects prevented the exploration of this field until the last third of the 20^{th} century. Only in the 1960's, advancements in experimental particle physics enabled physicists to begin building detectors sensitive to high-energy cosmic radiation. Currently the field is still in rapid development and new discoveries are made on an almost monthly basis as more sensitive detectors become operational. In this chapter, we attempt to give an overview of the field of high-energy astrophysics, beginning with a look at the historical background, followed by a brief description of the different branches that make up this field, and their experimental methods. In the case of gamma-ray astronomy, the different existing detection techniques are described in more detail, and an outlook on the future of detector development is given.

1.1 Historical perspective

From the very beginning of human civilisation, the science of astronomy played an important role in the development of early cultures. From the first astronomical observations recorded around 4000 B.C. in Egypt until the Renaissance age of Copernicus, Galileo and Kepler in the 16^{th} century, observing and analysing the motions of celestial bodies

was used, amongst others, for time keeping, navigation, and the prediction of astronomical events such as lunar eclipses. The idea of a heliocentric universe was only one of several theories born from these efforts, initially put forward by Greek astronomers and philosophers a few hundred years before Christ. The ground breaking discoveries that were made in the Renaissance and later up until the early 20^{th} century, both by astronomical observations and laboratory experiments, truly revolutionised our view of the universe. One of the discoveries affecting astronomy the most was the understanding that visible light is but a small part of the whole electromagnetic spectrum, covering a very narrow range in frequency and hence energy. But while the wave character of visible light was known for quite some time, leading to important applications such as spectral analysis of star light, observations of celestial objects remained restricted to detecting electromagnetic radiation in that very range of the spectrum using conventional optical telescopes for a long time. This finally changed in the early 20^{th} century, when technology had advanced to the point where infrared images of astronomical objects could be recorded. The first radio telescopes were constructed in the 1940's, benefiting from progress made in radio technology during WW II. Telescopes that were sensitive to ultraviolet (UV), cosmic microwave background radiation (CMBR) and X-rays followed within the next decades. Being able to detect electromagnetic radiation from each of these different energy bands greatly increased our understanding of the objects found in the visible universe. At the same time, it also resulted in the detection of hitherto unknown, increasingly bizarre classes of astrophysical objects. The corresponding need to find explanations for the existence of these objects and models for the physical processes involved, which are often not reproducible on Earth due to the extreme conditions present in these objects, raised even more questions, requiring ever more sensitive observations in all accessible wavelength regimes, and beyond.

The first publication addressing the possibility that gamma rays of tens to hundreds MeV energy may be produced in high-energy processes within certain celestial sources, along with *cosmic rays*, can probably be credited to Morrison (1958). Yet the expected gamma-ray fluxes were overestimated, and although several ground based gamma-ray[1] detectors were operational as early as 1960, their low sensitivities and lack of effective cosmic-ray background discrimination capabilities (see Chapter 3) prevented a statistically sound detection of any sources. It was left to the satellite-borne detectors SAS-2 and COS-B in the 1970's to detect the first statistically significant signals of gamma rays between 35 MeV and 5 GeV from an unexpectedly large number of sources (Ramana

[1]The transition between X-rays and gamma rays is usually defined to occur somewhere between 500 keV and 1 MeV.

Murthy & Wolfendale, 1993). The first *ground-based* gamma-ray detection was claimed by the Crimean Observatory, an atmospheric Cherenkov telescope operating in the TeV energy range, which in 1979 observed periodic emission from the pulsar Cygnus X-3 (Neshpor *et al.*, 1979). Unfortunately, many of the early gamma-ray detections could not be confirmed, and the field of gamma-ray astronomy had to go through a period of scepticism before experiencing a second boom following the successes of the improved detectors that became operational around 1990, most notably EGRET and Whipple (see Sec. 1.5). For a detailed review of the beginnings of gamma-ray astronomy, the reader is referred to Weekes (1988).

Despite the new generation of gamma-ray detectors which have been commissioned in the past five years or are expected to become operational within the next few years, the region of high-energy gamma rays remains and will remain poorly explored when compared to other wavelengths, mainly because of the relatively lower sensitivity of gamma-ray telescopes, but also due to the extremely large range of energies to be covered and the resulting necessity to resort to more than one method of detection.

Of course the field of high-energy astrophysics is not restricted to the detection of electromagnetic radiation in the form of X-rays and gamma rays. The far more abundant cosmic rays, which are highly energetic charged nuclei as well as electrons, were first discovered by Victor Hess in 1912, who ascended in a balloon up to heights of 5 km and noticed that his electroscope discharged more rapidly as the altitude increased. The origin of cosmic rays, particularly those of the highest observed energies (up to 10^{21} eV), remains a largely unresolved question. Cosmic ray detectors, usually extended arrays of scintillator or water Cherenkov detectors as well as atmospheric nitrogen fluorescence detectors, have been in use for several decades. Their sensitivity is directly coupled to their size, which also determines the maximum detectable particle energy, and so larger detectors are being built. Last but not least, another important type of cosmic messenger is the neutrino. Due to their extremely low interaction cross-section, the detection of high-energy cosmic neutrinos is still in its infancy, as several prototype detectors are currently being expanded. Cosmic-ray and neutrino astronomy will be addressed in Sections 1.3 and 1.4.

1.2 The high-energy universe

There are two fundamental differences between conventional and high-energy astrophysics: one is the mechanisms involved in the production of the emitted radiation,

and the other concerns the techniques employed to detect the latter.[2] In radio, optical or infrared astronomy, most of the detected radiation is produced via *thermal* processes, meaning that it is emitted as a result of the thermal motion of atoms, ions and especially electrons. Stars are prime examples of bodies emitting thermal radiation; their surface temperatures are generally of the order of several thousand degrees Kelvin, resulting in emission that is peaked in the visible range of the electromagnetic spectrum (around 1 eV). An object which radiates *only* thermally and does not reflect any light from surrounding sources is called a *blackbody*. The frequency distribution of the emitted intensity, and thus energy density, of a perfect blackbody depends only on the temperature of the emitting plasma, and is described by Planck's law of blackbody radiation:

$$I(\nu, T) \;=\; \frac{2h\nu^3}{c^2}\,\frac{1}{e^{h\nu/kT}-1} \tag{1.1}$$

Here I is the spectral radiance in units of [J s^{-1} m^{-2} sr^{-1} Hz^{-1}], T the temperature of the blackbody in Kelvin, ν the frequency of the emitted light, h and k the Planck and Boltzmann constants and c the speed of light. When comparing the frequency distributions according to Eq. 1.1 for different temperatures, as is illustrated in Figure 1.1, it can be seen that the wavelength of maximum emission λ_{max} is inversely proportional to the temperature:

$$\lambda_{\max} \;=\; \frac{b}{T} \quad\Rightarrow\quad T \;=\; 2337 \cdot \frac{E_{\max}}{\mathrm{eV}}\ \mathrm{K} \tag{1.2}$$

where b is *Wien's displacement constant* ($2.8978 \cdot 10^{-3}$ K m). Thus the peak energy $E_{\max}$ of the emitted radiation is directly proportional to the blackbody temperature. Thermal radiation emitted at X-ray energies, *i.e.* with energies in the order of keV, requires temperatures of millions of degrees Kelvin. Only a few objects in the universe are known to have such high temperatures, among them the Sun's corona. The majority of the radiation with X-ray energies and all of the gamma rays reaching Earth from space are therefore thought to originate from *non-thermal* processes.[3]

Currently, most of the proposed production mechanisms of non-thermal radiation in the universe are based on strong relativistically propagating magnetic fields, often containing *magnetic shocks*, capable of accelerating charged particles to ultra-relativistic energies. In these magnetic fields, the accelerated particles are then thought to produce UV or X-ray photons in the form of synchrotron radiation. They can also interact with ambient photons or nuclei to produce gamma rays with energies from a few MeV to tens

[2] For the purpose of this argument, the definition of the term *radiation* shall be relaxed to include massive particles and neutrinos, as opposed to only electromagnetic radiation (*i.e.* photons).

[3] This includes per definition also electromagnetic radiation due to fission and fusion.

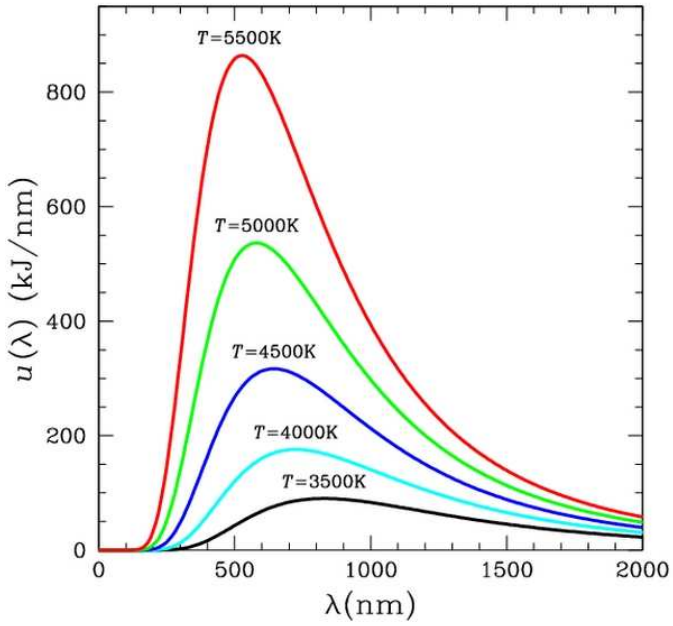

Figure 1.1: Energy density as a function of wavelength for perfect blackbodies of several different temperatures, according to Planck's law. The frequency of maximum emission is shifting towards higher values with increasing temperature (Wien's law).

of TeV. The most established sources of such non-thermal radiation include supernova explosions as well as the resulting neutron stars, pulsars and expanding nebulae (also called *supernova remnants* or SNR), *Gamma-Ray Bursts* (GRB), *Active Galactic Nuclei* (AGN), microquasars and X-ray binaries.

The origins of Gamma-Ray Bursts, brief bright flashes of X-rays and gamma rays from seemingly random directions in the sky, have remained a mystery for several decades. Only recently theories have been developed that associate GRB with highly beamed ejecta from supernova explosions (see *e.g.* Della Valle, 2005; Dar, 2005) or mergers of compact objects such as neutron stars or *black holes* (King, 2006). From observational evidence, it appears that the former mechanism is responsible for the majority of the long-duration GRBs (decay time $> 2\,\mathrm{s}$), whereas a second sub-class of intrinsically dimmer GRBs with emission lasting typically less than a second ("short bursts") may be caused exclusively by neutron-star and/or black-hole mergers. Active Galactic Nuclei are located, as the name suggests, in the centres of galaxies which emit strongly beamed plasma in the form of relativistic jets, making some of them appear unusually bright to the observer on Earth. A well known sub-class of this type of object are the so-called *quasars*. The core of an AGN is thought to be a *supermassive* black hole,

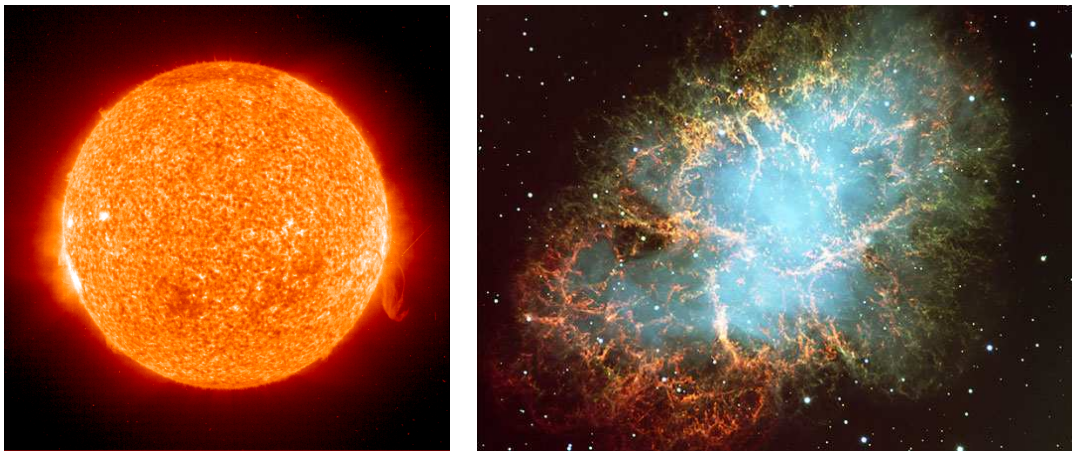

Figure 1.2: The Sun at visible to ultraviolet (*left*) and the Crab Nebula at visible wavelengths, making use of colour-coding to indicate different wavelength intervals (*right*). While the emission from stars such as the Sun, resulting from hydrogen and helium heated by nuclear fusion in the core, is almost entirely thermal, supernova remnants like the Crab Nebula feature relativistically expanding shells of gas which carry with them strong magnetic fields, capable of accelerating charged particles to very high energies. X-ray and gamma-ray emission is produced by the interaction of the latter with magnetic fields and ambient photons, respectively. Credits to SOHO (NASA) and Appenzeller (1998).

which – as opposed to ordinary (stellar-mass) black holes – boasts masses in the range of 10^7 to 10^{10} solar masses ($M_\odot$). The gravitational potential of the supermassive black hole forces ambient matter to form a rotating *accretion disk* around it, creating a very high spinning magnetic field. The strong magnetic field is responsible for the formation of bipolar relativistic jets emanating perpendicularly to the accretion disk, accelerating charged particles through co-moving magnetic shocks. These objects, as well as the various gamma-ray production models, will be examined in detail in Chapter 2. Microquasars and X-ray binaries are binary systems in which an expanding star is shedding mass onto a compact companion. X-ray and often gamma-ray emission is powered by the gravitational energy released in this process. In the case of microquasars, the compact stellar object is believed to be a black hole or neutron star, and gamma rays are produced by acceleration of charged particles in bipolar jets much like in AGN.

The second conceptual difference between conventional and high-energy astrophysics lies in the technique of detection. While radio, infrared and optical telescopes exploit the wave character of electromagnetic radiation by using the properties of reflection, refraction and interference to map out angular space, the wavelengths of X-rays and gamma rays are so small that none of these techniques can be applied. At these microscopic

wavelengths, the wave packets of electromagnetic radiation are no longer influenced by molecular or atomic structures, but interact directly with their constituents, *i.e.* the electrons, nuclei, nucleons, and at sufficiently high gamma-ray energies even quarks. As a result, photons with energies in the hard X-ray or gamma-ray domain cannot be focussed and refracted, as they penetrate all conventional focusing devices and react randomly with the constituent particles, nor can they be detected with the usual means. Instead, given that both high-energy astrophysics and particle physics experiments deal with detecting high-energy particles, it is not surprising that the same detection techniques are used in both cases. The field of experimental high-energy astrophysics is therefore often referred to as *astro-particle physics* (or *particle astrophysics*).

1.3 Cosmic Ray Astrophysics

As mentioned in Sec. 1.1, cosmic rays were discovered in 1912 as the first messengers of the universe's high-energy phenomena that were detected on Earth. Nevertheless, it took nearly half a century until they were studied in more detail. Cosmic rays are charged particles accelerated to very high energies by astrophysical sources located somewhere beyond the Earth's atmosphere. Due to the close connection between cosmic rays and gamma rays, both in production and in detection (cosmic rays represent the dominant background for gamma-ray detectors), this section is meant to give a brief introduction to the field of cosmic-ray astrophysics.

1.3.1 Cosmic ray production sites

Although the origin of cosmic rays is not yet known with certainty – at least not at the highest energies – and consequently the details of cosmic-ray acceleration are still under debate, the most likely scenarios include relativistically moving or spinning magnetic fields, which are thought to be present in galactic sources like supernovae, supernova remnants and pulsars, and in extragalactic ones such as AGN and GRB. Among these, the galactic sources with their slightly lower energetic scales appear to be good candidates for the production of cosmic rays up to energies of about 10^{18} eV, while cosmic rays at the very high end of the detected energy spectrum – commonly referred to as *ultra-high energy cosmic rays* (UHECR) – likely come from extragalactic sources (see also Sec. 1.3.2). Naturally other, yet unknown, production sites may also play a role, among them physical processes beyond the standard model, such as decays of supersymmetric particles and *primordial* black holes.

Theorists have developed several models which attempt to explain the emission of *very-high energy* (VHE)[4] gamma rays in objects such as the ones listed above. One popular model links the mechanism of cosmic-ray acceleration with the production of VHE gamma rays. (Various gamma-ray emission models for AGN will be examined in Chapter 2.) Due to the existence of galactic ($\approx 3 \cdot 10^{-6}$ G) and (much weaker) extra-galactic magnetic fields, the arrival directions of cosmic rays, which carry electric charge, do not point at their sources.[5] Instead, they are isotropic above a certain low energy limit, above which the local influence of the Sun is negligible. In contrast to cosmic rays, gamma rays are not deflected by the cosmic magnetic fields and therefore do point at their origins, provided they are not absorbed and re-emitted by background radiation on the way to the observer (Sec. 2.5). One of the chief motivations for gamma-ray astronomy is therefore the pinpointing of cosmic-ray acceleration sites. Very recently, the *Pierre Auger* cosmic-ray observatory in Argentina found indications for a non-statistical clustering of the highest-energy ($E > 6 \cdot 10^{19}$ eV) cosmic-ray events within 6° of the directions of nearby AGN (Auger Collaboration, 2008).[6] Previously, no excess of cosmic rays from the direction of any known X-ray or gamma-ray source had been unambiguously observed.[7] The Auger detector (see Sec. 1.3.4) with its enhanced sensitivity to the highest energy cosmic rays should in the near future be able to pinpoint point sources of UHE cosmic rays with improved statistical power.

1.3.2 The cosmic-ray spectrum

The observed all-particle energy spectrum is shown in Figure 1.3; the energy range over which primary cosmic rays have been detected ranges from 10^8 to about 10^{21} eV and is obviously very large. The flux density decreases steadily with energy according to a broken power law. Below 10 GeV, the cosmic-ray flux is modulated by the 11-year solar cycle, as the particle flux from the Sun is responsible for "absorbing" a fraction of the primary cosmic rays through pion production or photodisintegration. Due to this

[4] In the most common definition, the VHE range covers gamma-ray energies from about 10 GeV to tens of TeV. Gamma-ray energies above ~ 100 TeV are generally referred to as *ultra-high energies* (UHE).

[5] An exception, within a few degrees, are the extremely rare cosmic rays of the highest observed energies (above $\sim 10^{19}$ eV).

[6] If originating from these AGN closer than 75 Mpc, UHE cosmic rays of the mentioned energies are expected to be only minimally deflected by the intergalactic magnetic field.

[7] Claims of an excess of cosmic rays from the direction of the galactic centre reported by the AGASA and SUGAR collaborations could not be confirmed by recent data from the Auger Observatory (Abraham *et al.*, 2006). A point-source origin is only one of several explanations for the hints of a clustering (doublets and triplets) of UHECR events in the data of the AGASA and Yakutsk arrays (Tinyakov , 2001).

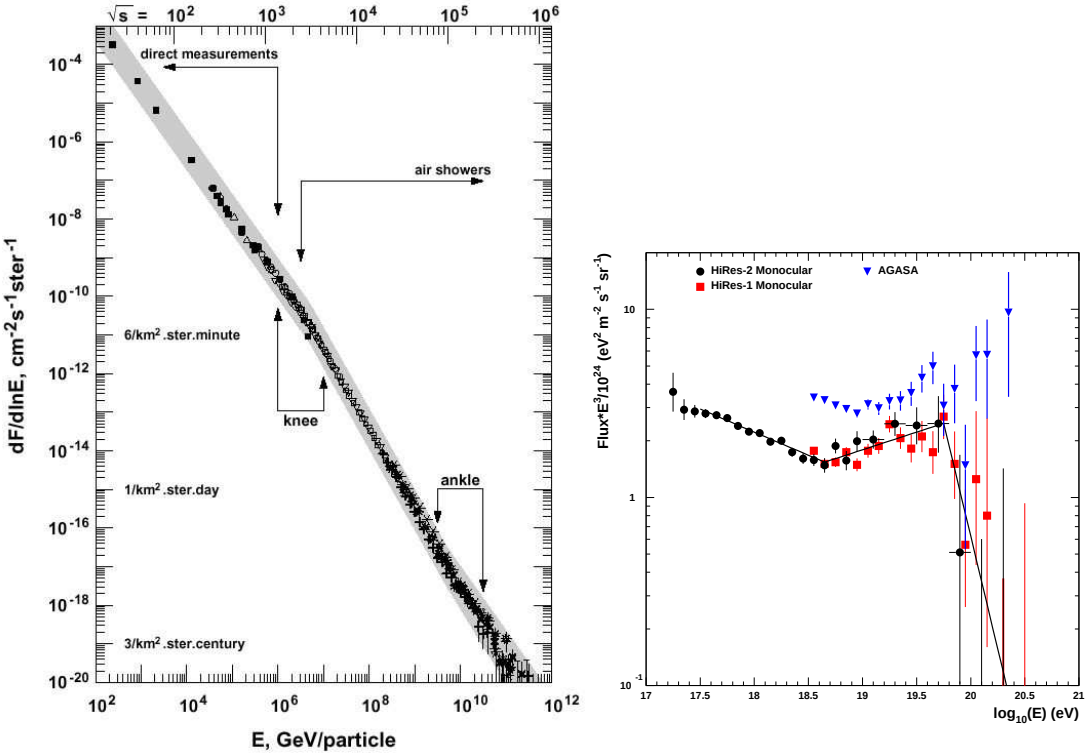

Figure 1.3: The all-particle cosmic-ray flux: across all energies (*left*) and zoomed in on the high energy end (*right*). The fluxes on the right plot have been multiplied by E^3 in order to emphasise the spectral features. Different symbols represent the measurements by various experiments. The "knee" and "ankle", two distinct changes in the spectral index of the power law around $3 \cdot 10^{15}$ eV and $3 \cdot 10^{18}$ eV, are clearly visible. Note the large error bars at energies above 10^{19} eV caused by poor statistics. The solid line represents a fit of a broken power law to the *HiRes* data, tuned to the prediction of a model which includes the GZK cutoff. Figures taken from Thomson (2006) and Bergman (2007).

attenuation effect and the deflection in the solar magnetic field, the extrasolar cosmic-ray flux is negligible below 1 GeV, whereas the contribution from the Sun itself increases with decreasing energy. Solar flares, for instance, are able to accelerate charged particles to energies of several GeV.

The composition of the cosmic-ray flux, measured directly by space- and balloon-borne detectors up to around 10^{15} eV and indirectly above, was found to vary slightly with energy. At energies below $\sim 10^{15}$ eV, the various measurements suggest that 86% to 89% of the particles are protons, 10% to 11% are helium nuclei (α-particles), up to 2% are electrons and the remaining 1% to 1.5% are heavier nuclei (Schlickeiser, 2003; Perkins, 2003; Ferrari, 2006). At very high energies, the heavier components appear

to be further suppressed in favour of the light ones, and the highest energy cosmic rays are almost exclusively protons. The given ranges indicate the current experimental uncertainties, which are quite large, especially at very high energies where the low fluxes force us to rely on indirect measurements. Only very small fractions of antiparticles (e^+, $\bar{p}$) have been observed; these are thought to originate from interactions of primary cosmic rays with interstellar gas. While all elements of the periodic table occuring in the solar system have been found in cosmic rays, the overwhelming majority of the remaining flux of heavier elements is made up of C, N, O and Fe. With the exception of nuclei that are weakly bound (such as Li, Be, B and F) and are therefore not commonly found in stars, but can be produced through spallation of heavier elements in interstellar gas, the cosmic ray composition resembles roughly that of stellar matter. (Apart from the just mentioned deficit of certain elements inside stars, both abundance distributions show the same "odd/even" effect caused by the fact that nuclei with even Z/N are more strongly bound.) The composition analysis thus favours a quasi-stellar origin of cosmic rays (Longair, 1992).

The most remarkable feature in the all-particle cosmic-ray spectrum is probably the so-called "knee", a spectral softening at about $3 \cdot 10^{15}$ eV, where the spectral index of the power law changes from 2.7 below to 3.1 above the knee. A multitude of possible explanations for this feature have been suggested, the most important ones being a change in the acceleration mechanism at knee energies, and a leakage of cosmic rays from our galaxy due to a reduced confinement efficiency by galactic magnetic fields at energies above the knee (Kampert *et al.*, 2004). At around $3 \cdot 10^{18}$ eV, another change in slope occurs. The spectral flattening around this energy is commonly referred to as the "ankle", and theories on its origin are even less certain than in the case of the knee. The most favoured idea involves an increased influx of cosmic rays, which at these high energies are believed to be almost exclusively protons, from production sites outside of our own galaxy. There is observational evidence that the ankle actually consists of a so-called "second knee", a "dip", and a "bump". These features are thought to be caused by the attenuation of extragalactic proton fluxes through electron-positron pair production in the cosmic microwave background (2^{nd} knee and dip), and a pile-up of even higher-energy protons (bump) that have lost part of their energy via the GZK-effect, which will be explained in the next section 1.3.3 (Berezinsky, Gasisov & Grigorieva, 2005).

1.3.3 The GZK cutoff

The universe is filled with a relic radiation from the time when radiation first decoupled from matter, released when electrons first combined with nuclei at a mean time of 372 thousand years after the Big Bang, which is known as *Cosmic Microwave Background Radiation* (CMBR). The CMB spectrum is a thermal (blackbody) spectrum corresponding to the temperature of the universe at the time of combination, redshifted according to the expansion of the universe since that time. At present, the temperature of the CMB is 2.725 K and thus its blackbody spectrum peaks at around $\lambda = 1.9$ mm.

As they travel through the CMBR, cosmic rays from extragalactic sources experience some energy loss in the form of electron-positron pair production. At ultra-high energies – above a threshold energy of $6 \cdot 10^{19}$ eV – they are predicted to lose energy also through photopion-production on the CMB.[8] As shown in Figure 1.4, the cross-section for photopion-production rises sharply with proton energy starting at the threshold energy, and its value at 10^{20} eV translates into a mean free path of only 50 Mpc. As a consequence, no cosmic rays above that energy should be reaching us if the production sites of these ultra-high energy (UHE) particles are much further than 50 Mpc from Earth (Ferrari, 2006; Albuquerque & Smoot, 2006). There are no known accelerators in our own galaxy which would be able to boost cosmic rays to energies around or above this threshold energy. All of the most likely sites of UHE cosmic-ray acceleration, which include AGN and GRB, are located in other galaxies further than 100 Mpc from us (for comparison, the diameter of our galaxy is about 30 kpc). Assuming there are no sources of UHECR within $\sim$50 Mpc, a cutoff in the observed cosmic-ray spectrum is thus expected between 10^{20} and 10^{21} eV. This effect was first proposed by Kenneth Greisen, Vadem Kuzmin and Georgi Zatsepin in 1966 (Greisen, 1966; Zatsepin & Kuzmin, 1966), and is since referred to as the *GZK cutoff*.[9]

At the highest cosmic-ray energies – above 10^{20} eV – only few events have been detected so far, which is reflected in the large error bars on the right plot in Figure 1.3. Furthermore, two of the three major cosmic-ray experiments sensitive to events of such high energies, *AGASA* and *HiRes*, seem to disagree not only in so far as the absolute

[8] This energy is determined by the mass and width of the Δ-resonance and the energy of the CMB photons.

[9] Note that if there *are* sources of UHECR within $\sim$50 Mpc, as suggested by recent Auger results (Auger Collaboration, 2008), the GZK cutoff can be expected to be smeared out.

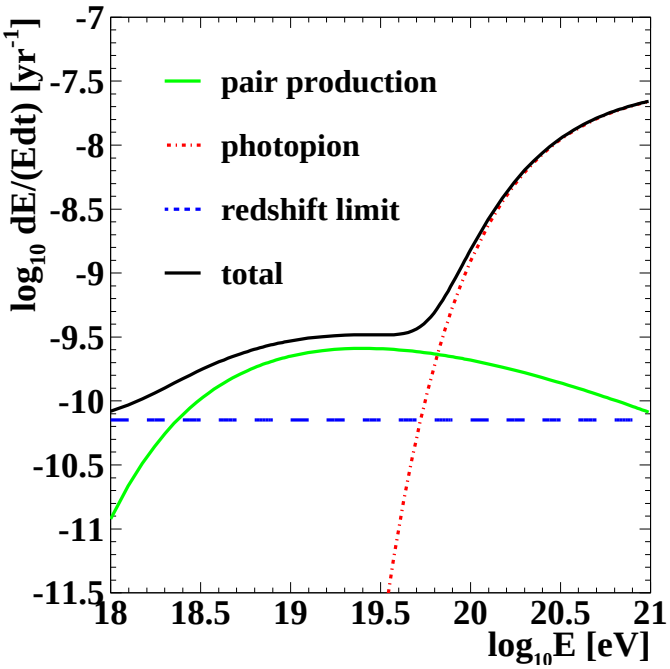

Figure 1.4: Double-logarithmic plot of energy loss as a function of energy, for a proton traversing intergalactic space at present epoch. Depicted are the dominating energy loss mechanisms of ultra-high energy cosmic rays, which are pair production ($p\,\gamma_{\mathrm{CMB}} \rightarrow p\,e^+\,e^-$) and photopion-production ($p\,\gamma_{\mathrm{CMB}} \rightarrow p\,\pi^0$ or $p\,\pi^+\,\pi^-$ or $n\,\pi^+$) on CMB photons. The redshift limit corresponds to the energy loss due to universal expansion. Adopted from Albuquerque & Smoot (2006).

energy scale is concerned[10], but also in regard to the question whether the GZK cutoff exists or not (see Fig. 1.3). Recently the HiRes collaboration has claimed to have enough statistical evidence to confirm the existence of the GZK cutoff (Thomson, 2006; Bergman, 2007), yet there is still some debate around this issue. The gigantic Pierre Auger Observatory in Argentina currently nearing completion (Sec. 1.3.4) is expected to be able to "see" enough cosmic rays of the highest energies to elucidate the shape of the spectrum in the vicinity of the GZK limit.

1.3.4 Detection technique

The design of cosmic-ray detectors varies depending on the targeted energy range. For low energies, at which the cosmic-ray flux density is sufficient to allow for small detector

[10] Berezinsky, Gasisov & Grigorieva (2005) shows that the sub-GZK parts of the spectra measured by both instruments can be brought into perfect agreement if their energy scales are re-calibrated using the spectral feature known as the *dip* (see Sec. 1.3.2).

areas, satellites or balloons are used. This method of direct detection has the advantage that it enables measuring of the primary flux unattenuated by the Earth's atmosphere. Very-high or ultra-high energy (*i.e.* TeV and up) cosmic rays are detected indirectly through the extensive air showers they produce in the atmosphere (the physics of air showers will be discussed in Chapter 3). The charged air shower particles are detected by means of large particle detector ground arrays, examples of which are the *KASKADE, Haverah Park, Tibet* (Tibet, 2005), *Yakutsk* (Egorova *et al.*, 2004) and *AGASA* (Takeda, 2003) experiments. More details on these types of detectors will be given in Sec. 1.5.3.

An alternative approach taken by the *HiRes* collaboration (see *e.g.* Westerhoff *et al.*, 2004) uses sensitive photodetectors to observe the faint fluorescence light produced by the spontaneous de-excitation of atmospheric nitrogen molecules following their excitation by highly energetic charged shower particles. The *Pierre Auger Observatory* (Mantsch, 2005), a cosmic-ray detector of unprecedented size which is nearing completion and is optimised for detecting UHE cosmic rays, combines these two complimentary techniques by operating 24 fluorescence telescopes overlooking a 3000 km^2 array of particle detectors made up of water Cherenkov tanks.

1.4 Neutrino Astrophysics

As of today, there has been no unambiguous detection of highly energetic neutrinos from astrophysical sources.[11] The main reason for the delay in the development of neutrino astronomy when compared to the neighbouring disciplines is the tiny reaction cross-section of neutrinos. Detecting the small neutrino fluxes expected from distant cosmic sources requires huge detector volumes, the construction of which is not only expensive but also technically challenging. There are several reasons, however, why the detection of energetic neutrinos from a cosmic origin would be truly ground breaking, and warrants the efforts which are currently being put into developing kilometre-scale underwater and under-ice neutrino detectors.

Since any acceleration process of cosmic rays is expected to be accompanied by an abundance of neutrinos (see Sec. 2.3.2), the detection of energetic neutrinos from a given object would unambiguously flag it as a cosmic-ray accelerator, which, as a welcome spinoff, would also shed some light on the gamma-ray production mechanisms present. In other words, neutrinos, which are neither deflected by magnetic fields nor

[11] Here we do not consider the well established detection of neutrinos from the Sun and from supernova SN 1987A, as their energies are in the keV to MeV range and thus several orders of magnitude below the energies of interest in the context of VHE gamma-ray astronomy. Other claims of astronomical high-energy neutrino detections lack a solid foundation.

attenuated by cosmic background radiation, can point us to the places where cosmic rays are accelerated. Another reason why ultra-energetic neutrinos are interesting is dark matter. Weakly interacting massive particles (*WIMP*s) such as the proposed *neutralino*, a supersymmetric non-baryonic particle which is regarded as the most promising dark matter candidate, would be clustered around gravitational attractors, even within the solar system. Any excess of high-energy neutrinos from the centres of the Earth or Sun would suggest the presence of such particles, since the decay of their annihilation products would produce neutrinos as well as gamma rays.

The main UHE neutrino detectors currently under development are *ICECUBE* (ICECUBE, 2006) at the South Pole, *ANTARES* (ANTARES, 2005) in a deep part of the Mediterranean Sea, and *BAIKAL* (see *e.g.* Wischnewski, 2006) in Siberia's Lake Baikal. All three experiments are designed to detect the Cherenkov light emitted by muons created or electromagnetic cascades initiated by neutrino interactions with the detector medium, *i.e.* ice or water (or, in rare instances, the underlying rock). Since there is an overwhelming background of downward going muons from extensive air showers produced by cosmic-ray interactions in the atmosphere above the detectors (Sec. 3.1.1), the trigger logic is such that *upward* moving Cherenkov light patterns are selected; in other words, the Earth is used as a radiation shield for atmospheric muons.

1.5 Gamma-ray Astrophysics

In this section we will briefly summarise the answers scientists hope to obtain from astrophysical gamma rays, and review the past, present and future experimental developments in the field of gamma-ray astronomy. Since the energy range covered by gamma rays spans many decades (from about 10^6 to above 10^{15} eV), more than one detection technique is required to cover the whole spectrum. The three main experimental approaches to gamma-ray detection will be described, along with their advantages and disadvantages.

1.5.1 Why gamma-ray astronomy?

As mentioned in Sec. 1.1, the vast gamma-ray band represents the least explored wavelength range of the electromagnetic spectrum. For several decades, the existence of highly energetic processes in the universe was being hinted at by cosmic rays arriving from random directions, yet before the advent of the first successful X-ray and gamma-ray detectors, the sites where these violent processes take place could only be inferred

from the detection of low-energy (radio and optical) electromagnetic radiation produced by the interactions of high-energy particles with surrounding matter and magnetic fields. Providing actual evidence for the existence of non-thermal processes in the universe was left to X-ray and gamma-ray astronomy. The three fundamental questions that gamma-ray astronomy is trying to answer are: what types of astrophysical objects emit at energies in the gamma-ray band, what do their energy spectra look like, and finally, of what nature is the time dependence of the emission? For extended objects, the morphology of the emission is also of interest. The answers gamma-ray astronomy can provide us with in regard to any or all of these questions help us to better understand the high-energy phenomena in the universe; in particular, they give us valuable clues about

- how particles are accelerated to extremely high energies within a given class of object,
- whether cosmic rays are accelerated in the same objects as part of the gamma-ray production process,
- how the laws of physics behave under extreme conditions such as intense magnetic and gravitational fields,
- the existence and nature of cold dark matter, and
- the spatial and spectral distribution of the Extragalactic Background Radiation.

Most of the above points will be discussed in more detail in Chapter 2.

1.5.2 Satellite-based detectors

The Earth's atmosphere is opaque to all electromagnetic radiation of frequency higher than UV-A (the longest wavelength UV light), in particular to X-rays and gamma rays. For gamma rays of GeV to TeV energies, the depth of the atmosphere corresponds to about 28 radiation lengths at sea level, and still about 16 radiation lengths at high mountain altitudes. It follows that gamma rays can only be observed directly from outside the atmosphere. Satellite-based gamma-ray detectors are the obvious solution.

The principal elements which satellite-borne detectors for gamma rays above 30 MeV consist of have not changed since the very first satellites. This is not surprising, given that the same concept is still widely used in particle physics detectors. At energies above 30 MeV, photons lose energy mainly due to pair-production. Therefore, the main body of a satellite detector consists of several layers of a high-Z *converter* material such as lead, with thickness of about one radiation length. Gamma rays interact with the converter

nuclei to produce electron-positron pairs. In this process, the e^-/e^+ pair retains (to sufficiently high precision) the energy and directionality of the original gamma ray, due to the much larger mass of the recoil nucleus. Located in between the converter layers is a tracking medium, which allows for the reconstruction of the paths of the e^-/e^+ pair and thus the primary gamma ray's direction. The tracking detector is traditionally a spark chamber, consisting of a series of metal plates put on high voltage and a space filled with an inert gas between them. Charged high-energy particles passing through the spark chamber ionise the atoms of the gas and create electric discharges between the metal plates, marking the particle's track.[12] Following the tracking device is a *calorimeter*, a detector used to measure the energy of the e^-/e^+ pair and hence that of the original gamma ray. This is most conveniently done with the help of a crystal scintillator (typically Na I or Cs I), which converts the pair's kinetic energy into a flash of visible or ultraviolet light. The light is then collected and converted to electric current using one or more photomultiplier tubes (PMTs). Last but not least, operating a space-based gamma ray detector requires a means of identifying and rejecting charged particles, since it is constantly exposed to a high flux of cosmic rays arriving isotropically from all directions. This is achieved with the help of an *anti-coincidence shield*, a detector which responds only to energetic charged particles but is transparent to gamma rays. A dome made of plastic scintillator (due to its faster response compared to crystal scintillator), surrounding the tracking detector and read out by PMTs, is commonly used.

There have been several space missions involving gamma-ray instruments aboard satellites in the late 20^{th} century. The first such satellite, *Explorer XI* (Kraushaar *et al.*, 1962), was launched in 1961 but its science program enjoyed limited success. In 1967, one of the U.S. Department of Defense's *Vela* satellites, designed to detect nuclear weapons testing on the ground and in space, provided the first serendipitous detection of a GRB (Klebesadel, Strong & Olson, 1973). Other space-based gamma-ray observatories worth mentioning are ESA's *COS-B* (Bignami *et al.*, 1975) and NASA's *SAS-2* (Fichtel *et al.*, 1975), which were very successful in the 1970's by observing the Milky Way's diffuse gamma-ray background and providing the first detections of discrete gamma-ray sources such as the Crab Nebula and Vela SNRs (Ramana Murthy & Wolfendale, 1993).

Without doubt the most successful space-based gamma-ray detector to date was the *Compton Gamma-Ray Observatory* (CGRO), a heavy (17-ton) satellite launched by NASA on April 5, 1991, a schematic diagram of which is shown in Figure 1.5. The CGRO

[12] In order to avoid saturating the gas with ions and electrons, the spark chamber plates are only put on high voltage whenever the other parts of the detector have generated a trigger signal. In spite of such measures, the spark chamber gas is used up after some time of operation and needs to be replenished periodically from an on-board supply.

satellite housed four separate, complementary gamma-ray detectors, each optimised for a different energy range and source type:

- *OSSE* (Oriented Scintillator Spectrometer Experiment) was intended to detect low-energy gamma rays (50 keV to 10 MeV) from steady sources.

- *COMPTEL* (imaging COMPton TELescope) was an experiment designed to produce a map of gamma rays in the 800 keV to 30 MeV range by means of Compton scattering.

- *BATSE* (Burst And Transient Source Experiment) consisted of eight independent NaI(Tl) crystal scintillation detectors, located at the corners of the satellite and monitored by PMTs. It served as an all-sky monitor for sources of transient X-rays and gamma rays in the energy range from 25 keV to 2 MeV, and was thus particularly suited for the detection of GRB. A sky map of the GRB detected by BATSE is shown in Figure 1.6, revealing for the first time the isotropic nature of these phenomena and thereby providing evidence for an extragalactic origin of GRB.

- *EGRET* (Energetic Gamma-Ray Experiment Telescope), sensitive to gamma rays between 20 MeV and 10 to 30 GeV with a peak sensitivity around 100 MeV, was very similar in design to previous gamma ray detectors such as SAS-2 and COS-B but with a much larger effective area (Thompson *et al.*, 1993). It consisted of a spark chamber divided into an array of slabs by interspersed converters made of tantalum sheets, a NaI(Tl) calorimeter and a scintillator dome as anti-coincidence shield (see Figure 1.7). The triggering was performed by two large layers of vertically stacked scintillator tiles which served as a time-of-flight system, selecting downward going electron-positron pairs which are not vetoed by a signal from the scintillator dome. The solid angle acceptance of the instrument was about 0.5 sr, the effective area rising to 0.12 m^2 at 1 GeV and decreasing at higher energies due to charged particle recoil from the calorimeter. As with all satellite detectors, the angular resolution improved with energy (from 3.5° at 100 MeV to 0.35° at 10 GeV), and the energy resolution (9 to 12%) decreased at high energies because of particle leakage from the calorimeter.

Covering the highest energy range, EGRET was also by far the most successful of the four CGRO instruments. It produced the first all-sky map of gamma rays above 100 MeV, which is shown in the top panel of Figure 1.8. Remarkable is the diffuse gamma-ray background from the galactic plane as well as several bright localised sources. Furthermore, it resulted in the detection of 271 gamma-ray point sources above 100 MeV,

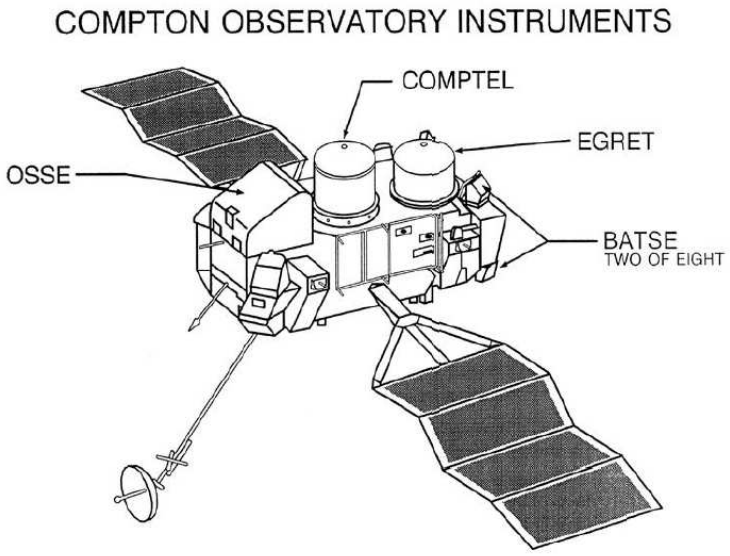

Figure 1.5: The Compton Gamma-Ray Observatory (CGRO) satellite and its instruments: Energetic Gamma-Ray Experiment Telescope (EGRET), imaging Compton Telescope (COMPTEL), Oriented Scintillator Spectrometer Experiment (OSSE) and Burst And Transient Source Experiment (BATSE). From www-CGRO (2006).

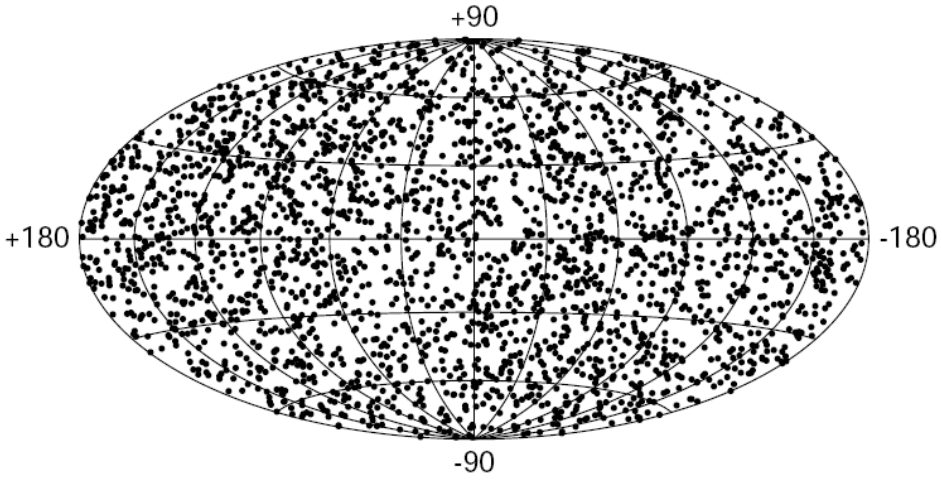

Figure 1.6: Sky map of the 2704 GRB observed by the BATSE detector during its lifetime, in galactic coordinates. From www-CGRO (2006).

including 66 confirmed and 27 possible AGN, 5 pulsars, 1 radio galaxy (Cen A), the Large Magellanic Cloud and one solar flare. It also detected gamma rays from 170 *unidentified* sources, most of them in the galactic plane, which could not yet be associated with counterparts at other wavelengths. These point sources have been published as the

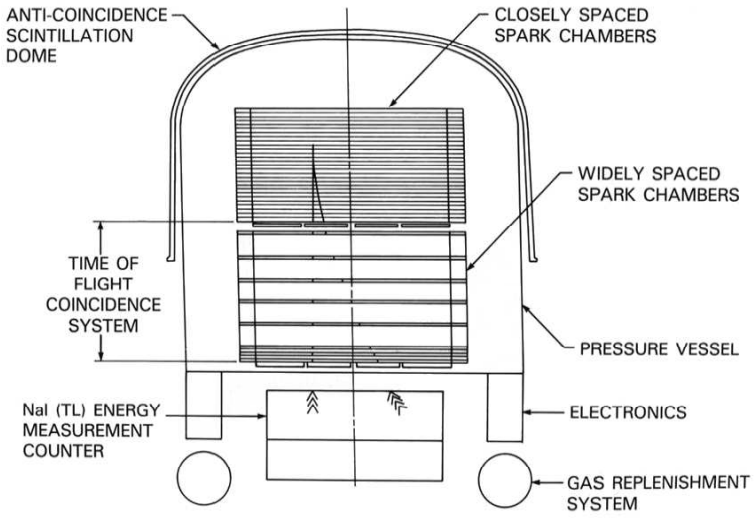

Figure 1.7: Schematic of the EGRET gamma-ray detector. From www-CGRO (2006).

Third EGRET catalog (Hartman *et al.*, 1999), and their coordinates are plotted in the bottom panel of Figure 1.8. CGRO was de-orbited on June 4, 2000. Nevertheless, the data it collected are still subjected to more sophisticated analysis techniques, and provide an invaluable reference for the planning and scientific interpretation of present and future gamma-ray observations.

1.5.3 Ground-based gamma-ray detectors

Although cosmic gamma rays themselves do not reach the surface of the Earth, satellite detectors are not the only kind of telescopes gamma ray astronomers have at their disposal. Gamma-rays of GeV energies or above can be detected from the ground by means of the secondary particles they produce when interacting with air. At such high energies, photons lose energy in matter exclusively through *pair production*. After travelling in average one radiation length in the atmosphere, the gamma-ray photon interacts with an air molecule of the upper atmosphere – between 10 and 15 km above sea level (ASL) – and produces an electron-positron pair. The electron and positron in turn lose a large fraction of their energy by producing bremsstrahlung photons in the fields of air nuclei. Subsequently, pair production and bremsstrahlung processes repeat themselves multiple times, producing an electromagnetic cascade consisting of hundreds of thousands to millions of charged particles and photons, which propagates down through the atmosphere. Particle cascades initiated by highly energetic cosmic particles are also called *extensive air showers* (EAS). The physics of air showers is described in more detail in Sec. 3.1.1.

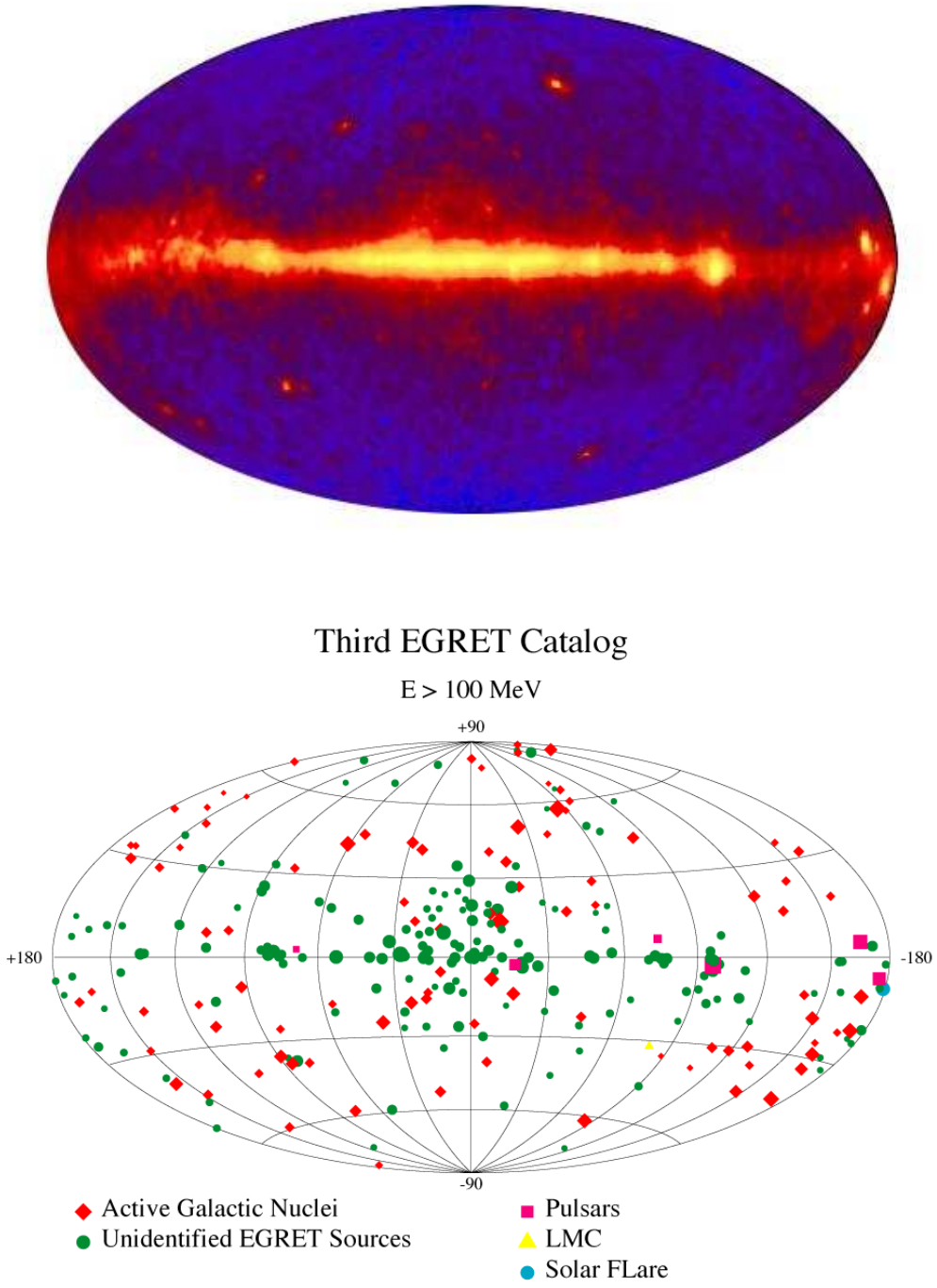

Figure 1.8: EGRET all-sky gamma-ray map (*top*) and the third EGRET catalogue of gamma-ray point-sources above 100 MeV (*bottom*), both in galactic coordinates. From www-CGRO (2006).

Experience from various experiments has shown that there are at least four ways in which extensive air showers can be detected from the ground:

1. The shower particles (for gamma-ray initiated showers these are electrons, positrons and low-energy gamma rays) can be detected directly using large particle detector arrays (Sec. 1.5.3).

2. The Cherenkov radiation produced by relativistic charged shower particles can be collected and focussed onto sensitive photodetectors (Sec. 1.5.3).

3. The UV to blueish light (300 - 400 nm wavelength) produced via fluorescence of atmospheric nitrogen molecules excited by energetic charged shower particles can be detected using the same technique as above (Westerhoff *et al.*, 2004).

4. Transient radio waves produced by the transverse electromagnetic field due to the separation of charge of the shower particles in the geomagnetic field can be picked up by sensitive radio receivers (Horneffer *et al.*, 2004).

Ground-based gamma-ray detectors have several disadvantages when compared to satellite detectors. The biggest challenge they face is the fact that charged cosmic rays interact with the atmosphere just like gamma rays, producing EAS with similar characteristics. (Fortunately, as we will see in Sec. 3.1.3, there are – although often very subtle – differences in morphology and timing between typical gamma-ray and cosmic-ray initiated air showers, without which ground-based gamma-ray astronomy would have to rely exclusively on statistical methods.) For energy thresholds[13] and acceptance angles of typical gamma-ray telescopes, the cosmic-ray flux exceeds the flux of gamma rays from even the brightest gamma-ray emitters by factors of several hundred.[14] Due to the overwhelming background of cosmic rays arriving isotropically from all directions, in practise only techniques 1 and 2 are of use for gamma-ray detection, while inherent difficulties in identifying gamma rays make the other two methods suitable only for cosmic-ray detection. Another disadvantage of ground-based gamma-ray astronomy is the fact that the detectors cannot be calibrated under controlled conditions with gamma-ray beams of known energy and luminosity at accelerator facilities, unlike satellite instruments. As

[13]The term *energy threshold* describes the turn-on energy of a detector. Since actual detectors do not have sensitivity functions that rise instantaneously from 0% to 100% at the energy threshold, the latter is usually defined as the peak of the *sensitivity curve*, which is the convolution of the detector effective area with the steeply falling power law of a typical gamma-ray source. This also corresponds to the energy at which the trigger rate of the detector is maximised. See Sec. 3.2 for details.

[14]When comparing the cosmic-ray flux within a solid angle of half-angle 0.5° to the gamma-ray flux from the Crab Nebula, the integral flux above 1 TeV is about 400 times higher for cosmic rays than for gamma rays (Ong, 1998).

a consequence, there is no easy way to directly measure the sensitivity and response of gamma-ray telescopes, and interpretation of the data has to rely heavily on detector simulations, which is subject to systematic errors. The fact that the atmosphere is an integral part of a ground-based detector adds another source of uncertainty: more so than for astronomy at other wavelengths, changes in atmospheric pressure, temperature, humidity and aerosol content pose a problem for gamma-ray astronomy, as they affect properties such as air density and transparency. The sensitivity and energy response of ground-based gamma-ray telescopes depend critically on these quantities, especially in the case of Atmospheric Cherenkov Telescopes (see next section). Therefore having a means of monitoring atmospheric conditions is important for ground-based detectors, not only for calibrating the detector in terms of absolute sensitivity, but also to provide a relative sensitivity scale, as fluctuating atmospheric conditions may mimic or obscure a variability of the source flux.

But the ground-based approach to gamma-ray detection has advantages over direct detection with satellites as well. For instance, since it is sufficient for identifying a gamma-ray initiated air shower to sample only a fraction of the lateral particle distribution on the ground, which is usually a circular or elliptical area from 260 m (in the case of Cherenkov light detectors) to several hundred metres in diameter, the effective area of the detector is no longer defined by its physical area, but by the area on the ground that is sampled with detector elements (or, for single imaging Cherenkov telescopes, by the size of the Cherenkov light pool; see Sec. 3.2.1). As a result, ground-based gamma-ray detectors can achieve very high effective areas by spreading a number of detectors over a sufficiently large area. Compared to satellite detectors, they are also quite inexpensive, and maintenance or possible expansion are much easier.

Satellite- and ground-based gamma-ray detectors complement each other conveniently: due to the rapidly falling emission spectra, which are usually governed by power laws, gamma-ray fluxes above about 30 to 100 GeV are too low to be detected by satellite detectors in statistically sufficient numbers, because of their restricted effective areas. On the other hand, gamma rays above several tens of GeV produce large enough cascades in the atmosphere to be observed by ground-based Cherenkov telescopes, while atmospheric absorption renders the detection of lower energy air showers from the ground impossible.

Atmospheric Cherenkov Telescopes

The most successful and promising technique for the detection of VHE gamma rays on the ground is without doubt the one employed by the *atmospheric Cherenkov telescopes*

(ACT). The physics of atmospheric Cherenkov light emission by EAS and the equipment used to detect it will be presented in Chapter 3. Here we would merely like to point out how these detectors relate to the other kinds of gamma-ray detectors, and what their strengths and weaknesses are.

As already suggested by the name, atmospheric Cherenkov telescopes detect the Cherenkov light emitted by highly relativistic charged air shower particles – for gamma-ray initiated air showers, these are exclusively electrons and positrons – in the atmosphere. One of the advantages of ACT is the fact that the Cherenkov light density on the ground scales with the energy of the primary gamma ray. Therefore, the atmosphere serves as a giant calorimeter for the ACT, *albeit* one whose output depends very sensitively on the prevailing atmospheric conditions (see previous section). With the most recent progress made in ACT development, detection energy thresholds as low as 50 GeV can be achieved (Mazin, 2006). At energies above about 50 TeV, *air shower arrays* are more suitable for gamma-ray detection, mainly due to the ACT's high low-energy event rates and degraded energy resolution for UHE gamma-ray showers. Another drawback of ACT is their short duty cycle compared to the other types of gamma-ray detectors; due to the high sensitivity of the photodetectors and the light-absorbing character of air-borne water molecules atmospheric Cherenkov light detection requires mostly clear, moonless nights. This results in a duty cycle of about 20% to 30%, whereas satellite detectors and air shower arrays can operate between 90% and 100% of the time.

The pioneer experiment in the field of atmospheric Cherenkov telescopes is the *Whipple* telescope on Mt. Hopkins in Arizona, USA (Kildea *et al.*, 2007). Consisting of a single dish-shaped optical reflector of 10 m diameter with a PMT camera at the focal point, Whipple is a so-called *imaging* atmospheric Cherenkov telescope.[15] Constructed in 1968, it was the first ground-based instrument to detect gamma rays from the Crab Nebula at high sensitivity (Weekes *et al.*, 1989) and with an energy threshold in the TeV range. The telescope was upgraded in the years following and is presently still operating, now with a much improved peak sensitivity of around 300 GeV. The second early imaging telescope worth mentioning is *CAT*, located in the French Pyrenées (Barrau *et al.*, 1998; Piron *et al.*, 2001). No longer in operation, CAT had a higher energy threshold due to its smaller size, but nevertheless made some important spectral measurements above 1 TeV (see *e.g.* Djannati-Ataï *et al.*, 2002). The first imaging ACT capable of making *stereoscopic* Cherenkov observations by using more than one imaging telescope in conjunction was *HEGRA* on La Palma, Canary Islands (Daum *et al.*, 1997). HEGRA

[15] Both the imaging and non-imaging technique for detecting atmospheric Cherenkov light showers will be described in detail in Sec. 3.2.

consisted of four (five from 1998 on) small reflectors of 3.4 m diameter, and produced many notable results until its disassembly in 2002 (HEGRA, 2003). Observing a source with an array of imagers has many advantages over using just a single telescope, as we will see in Sec. 3.2.1.

As of the time of writing, a new generation of imaging atmospheric Cherenkov telescope arrays is gradually becoming operational. *VERITAS* (Holder *et al.*, 2006) is under construction in Arizona, USA, with two out of four planned reflectors completed. *MAGIC* (Ferenc, 2005) on La Palma (Canary Islands) and *CANGAROO-III* (Kubo *et al.*, 2004) in Woomera (Australia) have recently been completed and are in the commissioning phase, producing first science results. Last but not least, the *High Energy Stereoscopic System* (HESS) array of four imaging telescopes (Hinton, 2004), located in the Khomas highland of Namibia and fully operational since 2004, has already detected gamma rays above 100 GeV from a substantial number of new sources and published upper limits on many others. The former include several point-sources in the galactic centre ridge (Aharonian *et al.*, 2006). HESS has also produced detailed VHE gamma-ray maps of several galactic extended objects. Technical aspects of the different ACT designs will be addressed in Sec. 3.2.

Figure 1.9 shows a sky map of all VHE gamma-ray sources detected as of September 2007. It shows 19 extragalactic sources, all of them AGN (16 HBL, one LBL, one FSRQ and one FR I radio galaxy), which are distributed more or less evenly across the sky. As expected from the galactic sources, of which there are currently 56, they are located predominantly in the galactic plane and there mostly in direction of the galactic centre. They include Pulsar Wind Nebulae (PWN), Supernova Remnants (SNR), X-ray binaries (microquasars), the Galactic Centre, an Open Cluster as well as many still unidentified sources. TeV emission from the following four previously reported sources could not be confirmed by recent, more sensitive observations with the HESS or CANGAROO-III arrays of imaging ACT, respectively: the Vela pulsar, SN 1006, NGC 253, and PSR 1706-443 (Ong, 2006).[16] It should be noted that conflicting observation results similar to the above are still not uncommon in VHE gamma-ray astronomy, which is mainly due to the extreme variability of some of the sources, but also because of rather complicated data analysis procedures required to correct for the cosmic-ray background, and the statistical nature of the technique. All of the plotted TeV gamma-ray emitters were discovered by imaging atmospheric Cherenkov telescopes, and most of them by the

[16] For all four objects, the earlier detection claims were made by the CANGAROO experiment. In the case of the starburst galaxy NGC 253, the CANGAROO collaboration has meanwhile retracted its claim (Itoh *et al.*, 2006).

new detectors HESS and MAGIC in the three years before publication of this map.[17] Both instruments are continuing to examine other potential VHE sources, and are also being upgraded. VERITAS and CANGAROO-III are starting to take data as well, and together with GLAST (see Sec. 1.5.4) "closing the gap" from the low-energy end (see Sec. 1.5.4), many more high-energy gamma-ray sources are expected to be detected in the coming years.

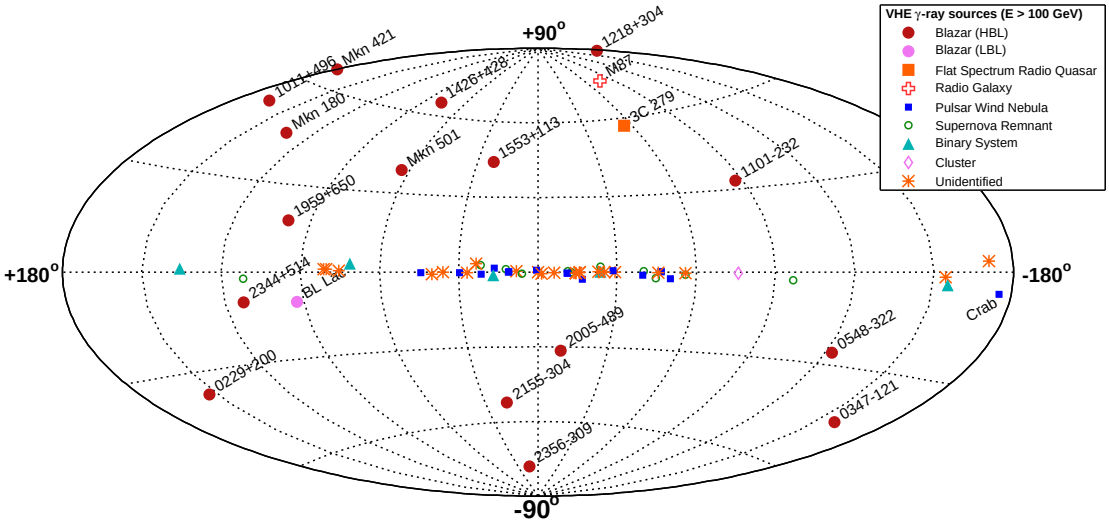

Figure 1.9: Sky map of detected VHE gamma-ray sources as of September 2007, in galactic coordinates. Only extragalactic sources and the Crab Nebula are labelled. The galactic sources cluster in the galactic plane towards the centre of the galaxy. Credits to www-Wagner (2007).

Air shower arrays

Air shower arrays, or *particle arrays*, were developed from cosmic-ray detectors and are often also used for cosmic-ray research. They consist of individual detector elements sensitive to the energetic charged particles of extensive air showers, spread out over relatively large areas. The individual detector elements are usually made of scintillators that are read out by PMTs, but designs consisting of Cherenkov light producing tanks filled with water are also in use. Needless to say, these types of detectors are only sensitive to gamma rays whose initial energy is high enough to allow a detectable number of energetic air shower particles to reach the ground level, and thus have higher energy thresholds than Cherenkov telescopes. Air shower arrays are of fixed geometry and

[17] Until 2000, the TeV catalogue contained no more than six VHE sources – two AGN and four pulsars/SNRs (Fortin, 2005). The lowest achievable energy threshold of ground-based detectors at the time was 250 GeV.

cannot be pointed, however their fields of view are several tens of degrees (although the energy threshold increases with increasing zenith angle), and thus large enough to permit observation of a large fraction of the sky. A definite advantage over ACT is the fact that they can be operated during the day and in any kind of weather, permitting a high duty cycle. Since the particles of an EAS are highly relativistic, they arrive on the ground in the form of a thin wavefront no thicker than about 1 m at the centre, which is perpendicular to the arrival direction of the primary particle. Therefore, good timing resolution of the detector elements (about 1 ns) allows for reconstruction of the primary gamma-ray direction. The energy of the primary is estimated by measuring the particle density as a function of the distance from the *shower core*, which is defined to be the intersection of the primary particle direction with the ground. The core location is approximated to be the centre-of-gravity of the lateral particle distribution.

The higher the energy of the primary particle, the deeper into the atmosphere the resulting EAS will penetrate (for more details, the reader is referred to Sec. 3.1.1). In order to keep the energy threshold as low as possible, the preferred sites for air shower arrays are therefore at high elevations. Since EAS due to gamma rays below about 100 GeV terminate before reaching the ground even at mountain altitudes, air shower arrays are not sensitive to gamma rays below 1 TeV. One of the first air shower array experiments dedicated to the detection of gamma rays was *CASA-MIA* (Borione *et al.*, 1994) in Utah, which operated in the 1990's and consisted of two separate interspersed arrays of scintillators. One of the arrays was buried underground and served as a muon detector, assisting in the rejection of background air showers from cosmic rays. Due to its high energy threshold of about 100 TeV (Ong, 1998) and limited sensitivity, CASA-MIA detected no sources, but placed upper limits on several source candidates. The *Tibet Air Shower Array* (Tibet, 2005), located on a high plateau in Tibet at an altitude of 4300 m ASL and thus the highest ground-based gamma-ray detector, is an array of 761 scintillator detectors of 0.5 m^2 size, covering an area of 37000 m^2. *Milagro* (Smith, 2005), a detector consisting of a large light-tight pool filled with water and arrays of PMTs at the bottom, uses the Cherenkov effect of charged air shower particles in water to detect EAS due to gamma rays above a few TeV. As in all ground array experiments, cosmic-ray rejection on an event-by-event basis is difficult, and only a search for localised excesses in the isotropic "sea" of cosmic rays can reveal the presence of a gamma-ray source. Both Milagro (Atkins *et al.*, 2004; Saz Parkinson *et al.*, 2004) and Tibet AS (Amenomori *et al.*, 2005), claiming energy thresholds as low as ~ 3 TeV, have recently conducted successful all-sky surveys of the northern hemisphere, and detected emission from the Crab Nebula and the flaring blazar Mrk 421 (the two brightest TeV sources in

the sky). Milagro also confirmed the detection of very extended TeV gamma-ray emission from the Cygnus arm of our own galaxy, thought to originate from galactic cosmic rays interacting with the dense gas clouds located in that area of the galaxy (Saz Parkinson *et al.*, 2004). In addition to the confirmed detections, both experiments accumulated measurable excesses from the directions of several other known TeV emitting sources, although with significances below $5\,\sigma$ above background and therefore not counted as "detections". Because of their large fields of view, air shower arrays are particularly well suited for all-sky surveys including the search for transient objects such as GRB, as well as for detecting gamma rays from diffuse objects which are too extended to be seen by ACT.

1.5.4 The future of gamma-ray astronomy

As mentioned previously, the field of gamma-ray astrophysics is rapidly being expanded, both in the experimental and theoretical areas. In concluding this chapter, we shall now take a brief look at ongoing and future developments in detector design and performance.

The unexplored window

As it should have become clear from the above sections, satellite- and ground-based gamma-ray telescopes complement each other in terms of the energy bands they cover. Nevertheless, at the time of writing there remains a poorly explored energy range where the two detection techniques fail to overlap, namely between about 10 and 300 GeV. The new imaging ACT (see Sec. 3.2.1) are gradually beginning to close the gap from the upper end, but until the availability of new and more sensitive satellite-based detectors, our knowledge of what happens in the gamma-ray spectrum between 10 and 100 GeV will likely remain rather patchy. Ironically, as we will see in Chapter 2, exactly this energy band is of particular interest to astrophysicists. Furthermore, as we can see in Figure 1.8, a myriad of cosmic sources have been found to emit gamma rays in the MeV band. At TeV energies, on the other hand, most of the sources seem to have disappeared (see Figure 1.9). While part of this "mystery" may be attributed to the fact that the emitted gamma-ray fluxes decrease with energy according to power laws, and may thus fall below the sensitivity thresholds of ground-based detectors, the physical processes of the emission and gamma-ray absorption effects during propagation are more likely to be responsible. Whatever the reason for the discrepancy in source populations, "closing the gap" between the two energy regimes is important if we wish to understand the nature of emission processes and what happens to the gamma rays on their way from the sources

to Earth. The expected performance of the new generation of gamma-ray telescopes, currently in the commissioning or testing phase and producing promising initial results, raises hopes that the energy gap is about to be bridged.

Upcoming satellite experiments

There are currently two major space-based gamma-ray observatories in orbit, which came online only recently and are expected to produce exciting results within the next few years. Both benefit from major advancements in high-energy physics detector development since the days when EGRET was launched.

The first new gamma-ray satellite that was commissioned is the Italian Space Agency's *Astro-rivelatore Gamma a Immagini LEggero* (AGILE; Pittori & Tavani, 2004), which was launched in April 2007. AGILE carries two instruments, one detector for hard X-rays and a gamma-ray telescope sensitive to photon energies from 30 MeV to 50 GeV. Silicon plane detectors are used for tracking, which results in better angular resolution than EGRET's, but with comparable sensitivity. Other improvements over EGRET include a much larger field of view ($\sim 3\,\mathrm{sr}$) and excellent timing capabilities, making it ideal for observing transient sources such as GRB and millisecond pulsars.

On June 11, 2008, NASA launched the much anticipated *Gamma-Ray Large Area Telescope* (GLAST), whose main instrument, the *Large Area Telescope* (LAT), is similar in design to EGRET but with a much better performance (see *e.g.* Ormes, 2003; Carson, 2006). Instead of the segmented spark chamber, LAT features 18 layers of high efficiency, fine pitch silicon strip detectors. The calorimeter consists of 1536 CsI(Tl) scintillator crystals in 8 layers. The anti-coincidence shield is made of 89 segmented plastic scintillator tiles. LAT has more than five times EGRET's effective area (8000 cm^2 compared to EGRET's 1500 cm^2), four times the field of view ($> 2\,\mathrm{sr}$ instead of $0.5\,\mathrm{sr}$), almost twice the angular resolution (3° at 100 MeV compared to 5.8° at 100 MeV and about 0.15° above 10 GeV) and 6 to 10 times more accurate source location capabilities. An increased sensitivity, which is expected to be about 17 times higher than that of EGRET, combined with the larger effective area will make LAT sensitive to gamma rays from 20 MeV to 300 GeV, although with very low statistics at the high-energy end.[18] GLAST also carries a burst monitor experiment (GBM) similar to BATSE. If they perform as expected, GLAST's instruments are bound to make an astounding number of new discoveries within the gamma-ray universe, and provide coverage of most or all of the unexplored energy window.

[18] At the present time it is difficult to find information about LAT's actual sensitivity above 100 GeV. At the current mission stage, publications are often quite optimistic.

Apart from these two dedicated gamma-ray satellites, the *AMS-02*, a cosmic-ray detector which is to be attached to the International Space Station, is also projected to have gamma-ray detection capabilities up to GeV energies (Gentile, 2007).

Ground based detectors

Construction has begun on the expansion of two of the existing new ACT, namely HESS and MAGIC. Both experiments will be upgraded to include one extra large Cherenkov dish: HESS is installing a huge reflector of 28 m diameter in the centre of the four existing telescopes (Vincent *et al.*, 2005), and MAGIC will receive a second telescope 17 m in diameter (Baixeras *et al.*, 2005). The MAGIC collaboration is also planning to replace their PMTs with new hybrid photodetectors (HPD), which have a quantum efficiency of nearly 50%, and install faster FADCs. In addition to the upgrades to existing Cherenkov telescopes, which are well under way, designs for completely new Cherenkov telescopes have been in discussion as well. The *MACE* project as proposed will consist of two ACT dishes of 21 m diameter at a high altitude site (Hanle, India at 4200 m ASL), and aims for a threshold well below 50 GeV (Koul *et al.*, 2005).

There is currently one new air shower array for gamma-ray detection in construction. The *ARGO-YBJ* air shower array at Yangbajing, Tibet (which is the same site as the Tibet AS array) will cover an area of 6400 m^2 (Di Girolamo *et al.*, 2005). The novelty of this detector is the fact that it uses resistive plate chambers (RPCs) for particle detection. About 50% of the detector is complete and operational, and first results have been presented. Other ideas include a further development of Milagro called *High Altitude Water Cherenkov Experiment* (HAWC). HAWC is proposed to consist of a square pond of side length 300 m with a large number of PMTs in a single layer at its bottom. It is to be located at a much higher altitude than Milagro (> 4000 m, compared to Milagro's 2650 m), which would allow for a similar energy threshold as that of the Whipple telescope but with 4000 times the exposure (Smith, 2005), paving the road for all-sky surveys and GRB detections at unprecedented low energies.

CHAPTER 2

ACTIVE GALACTIC NUCLEI

In this chapter we will take a closer look at the most violent known objects in the universe, *Active Galactic Nuclei* (AGN). After an introduction to the subject, a classification of the various types of AGN in accordance with current observational evidence will be presented. Subsequently, the relevant aspects of the production mechanisms for highly energetic gamma rays in these objects will be addressed, as well as the effects concerning the propagation of gamma rays through intergalactic space to the observer. In the end, the current state of knowledge about the blazar-type AGN H 1426+428 will be discussed.

2.1 Active Galaxies

The discovery of active galactic nuclei (AGN) is the result of observations at all astronomically accessible wavelengths of the electromagnetic spectrum, from radio waves to X-rays, and – since recently – also gamma rays. Due to the unavailability of appropriate radio and X-ray telescopes until the second half of the 20^{th} century, the history of the discovery of these energetic objects, as well as similar phenomena such as GRB, and of their theoretical treatment is very young, as is part of the underlying physical theory (general relativity, black hole physics). As the current state of AGN physics can be better put into perspective by knowing how this field developed, we shall begin this chapter with an outline of its history. Since black holes and their properties are (literally) central to AGN physics, a brief introduction to the former will be given in the beginning.

2.1.1 Black Holes

The development of a formalism that predicted the existence of *Black Holes* and described the laws of physics in and around these objects commenced in 1916, when Karl Schwarzschild[1] found the first solution of Albert Einstein's theory of General Relativity. Soon thereafter followed the prediction of a gravitational collapse of heavy stars into

[1] In an amusing coincidence, the German word "schwarz" means *black*, and "Schild" signifies *shield*.

objects whose strong gravitational fields even photons could not escape. The existence of a singularity at the precise location of the black hole was postulated by Roger Penrose and Stephen Hawking. The radius around a non-rotating (point-like) singularity below which no light can escape its gravitational pull is called *Schwarzschild radius*, the spherical surface described by that radius *event horizon*. Space outside the Schwarzschild radius is thus "shielded" from matter that has crossed the event horizon, *i.e.* there is no causal connection between the interior of the event horizon and the outside. The shielding effect of the event horizon can be explained by means of the *gravitational redshift*, which – due to the infinite curvature of space-time at that surface – approaches infinity at the event horizon. The wavelengths of photons coming from the event horizon are thus infinite, *i.e.* unobservable.

Up to then these exotic, not directly observable objects were commonly referred to as "frozen stars", due to the fact that time as measured by a distant observer is expected to slow down to a complete halt at the event horizon. This curious fact is prescribed by the space-time metric for the space outside the event horizon, which for the (simplest) case of a non-rotating black hole is the Schwarzschild metric. The term "black hole" was coined only in 1967 by John Archibald Wheeler (Bekenstein, 2005), who also introduced the "*no-hair theorem*", according to which a black hole can be described by no more than three parameters: mass, angular momentum and charge.

For the purpose of an introduction it suffices to point out that black holes are believed to be electrically neutral, but they (or rather, their event horizons) do come in different shapes and sizes. Since both stars and galaxy centres rotate, black holes resulting from the gravitational collapse of material from either of these objects are also expected to carry angular momentum. The event horizon of a black hole with non-zero angular momentum, also called *Kerr black hole*[2], will be an ellipsoid instead of a sphere, and the point singularity is replaced by a ring. There also exists a region around the event horizon – the so-called *ergosphere* – where space essentially rotates along with the black hole at a speed high enough that even particles travelling at the speed of light are unable to escape co-rotation.

Although by definition of infinitely small size, the *observable* physical properties of black holes are largely determined by the size of their event horizon, which in the simplified case of non-rotating black holes is proportional to the enclosed mass (see Eq. 2.2). The "radius" of a 10 solar-mass ($M_\odot$) black hole is thus about 30 km, that of a *supermassive* black hole with $M_{BH} = 10^8\, M_\odot$ about 300 million km. If Earth were

[2]In 1963, the mathematician Roy Kerr solved Einstein's field equations for the case of a rotating black hole. The Schwarzschild metric is in this case replaced by the more general *Kerr metric*.

somehow made to collapse to a black hole, its Schwarzschild radius would be slightly less than 9 mm. We will briefly return to the discussion of black holes in Sec. 2.2.2.

2.1.2 From Black Holes to Active Galactic Nuclei

The first observational evidence for the existence of an unusual class of objects which were later identified as *Active Galactic Nuclei* (AGN) was provided by radio astronomy. The very first discrete source of radio waves other than the Sun, which was also the first radio source outside of the Milky Way that could be associated with an optical counterpart, was found to be a galaxy in the constellation of Cygnus. This *radio galaxy*, henceforth named *Cygnus A*, appeared to have a very peculiar morphology: instead of one, there seemed to be two nuclei, laterally displaced on either side of the galaxy centre, and around the latter there appeared to be a ring of dust. Various theories about the origin of the strong radio emission (Cyg A is the brightest constant radio source outside of our own galaxy) were put forward, and one popular explanation involved the merger of massive galaxies. After the discovery of many more radio galaxies during the following decade, however, it became apparent that the true answer had to be related to the conspicuous optical jets and large dual *radio lobes* which were often seen in these types of objects.

In the 1960's, the search for more radio galaxies revealed another phenomenon: a population of strong radio sources associated with faint star-like objects, which were henceforth named *quasi-stellar radio sources*, or, in brief, *quasars*. Later observations using higher resolution telescopes revealed that these quasars are in fact not point sources, but consist of a bright core with a fuzzy halo around them. When their severely redshifted[3] broad emission line spectra were eventually associated with the Balmer lines of hydrogen, astrophysicists were puzzled by the inferred redshifts, which were much larger than expected. In order to match the observed brightnesses of quasars, their intrinsic bolometric luminosities would have to be in the range of 10^{45} to more than 10^{48} erg s^{-1} (Carrol & Ostlie, 1996). Quasars are thus several orders of magnitude brighter than normal galaxies, while their bright emission region is much smaller than the size of an average galaxy. This was not easily explainable by contemporary models, and led many astronomers to believe that the observed emission lines could not possibly originate from the quasars, or that new exotic physics might be responsible for the high luminosities.

[3]The *cosmological redshift* z (of the light emitted by an object) due to the Hubble expansion of the universe, defined as $z \equiv (\lambda_{\rm observed} - \lambda_{\rm emitted})/\lambda_{\rm emitted}$, is an observationally based measure of an object's distance.

In 1964, the theorists Edwin Salpeter and Yakov Zeldovich independently came up with the hypothesis that matter heated to very high temperatures by falling toward a black hole could result in a bright emission at optical and X-ray wavelengths. The first quantitative calculations of the expected emission spectrum were done by Donald Lynden-Bell and Martin Rees in 1969 and 1971, respectively. According to their model, gas from the companion star is "sucked in" by the gravitational field of the black hole, spiralling down towards the event horizon in the form of a viscous *accretion disk*, heating up to millions of degrees Kelvin in the process. The glow from the outer portion of the heated disk would produce the bright optical emission seen in quasars, while X-rays would be produced in the extremely hot inner regions of the disk. The confirmation of this idea came in 1971 when combined X-ray, optical and radio observations of the X-ray source *Cygnus X-1* provided strong evidence that the source's X-ray emission was due to accretion of matter from a bright blue star onto a compact object, which most likely was a black hole with the mass of about 10 solar masses. The discovery of this X-ray binary in our own galaxy led the way to an eventual acceptance of black holes as the central powerhouses of quasars, with the difference that the, by orders of magnitude, higher luminosities of the latter required the presence of much more massive black holes, so-called *supermassive* black holes, which were thought to be present at the centres of quasars and boast masses above $\sim 10^7\,\mathrm{M}_\odot$.[4]

Building on the idea of accreting black holes, Lynden-Bell, Rees and their colleague Roger Blandford subsequently showed that the swirling motion of the accretion disk material would strengthen any magnetic field present, creating two spirals of spinning magnetic field lines on either side of and perpendicular to the disk. These whirling magnetic fields would be able to accelerate and collimate jets of charged particles, and thus provided a natural explanation for the powerful jets seen to emanate from the centres of radio galaxies in optical and X-ray observations. A connection between radio galaxies and quasars was thus established, in that one and the same phenomenon could explain the unusual emission characteristics of both objects: an accreting supermassive black hole in the centre of a galaxy, in other words an *active galactic nucleus*. At present it is estimated that only a few percent of all observed galaxies are active.

As we will see in Sec. 2.4, radio galaxies and quasars represent only two manifestations of the AGN phenomenon. A general unified model of AGN will be introduced in the following Section 2.1.3.

[4]Today the estimated masses of supermassive black holes range from several 10^6 to about 10^{10} $\mathrm{M}_\odot$. According to Lyndon-Bell and Rees, an accretion of one solar mass per year would in this scenario suffice to produce the observed luminosities of quasars.

2.1.3 The AGN paradigm

Building on the basics of the current, generally accepted AGN model which was outlined in the previous section, we now describe the main components of this paradigm. It should be stressed that although widely popular, not all aspects of this model are currently experimentally *proven*, nor can it be excluded that alternative processes play a role. In order to conceptualise the various components of an AGN, a schematic diagram such as that shown in Fig. 2.1 may be useful. It should be noted that this diagram is not to scale, but all the length scales are approximately logarithmic.

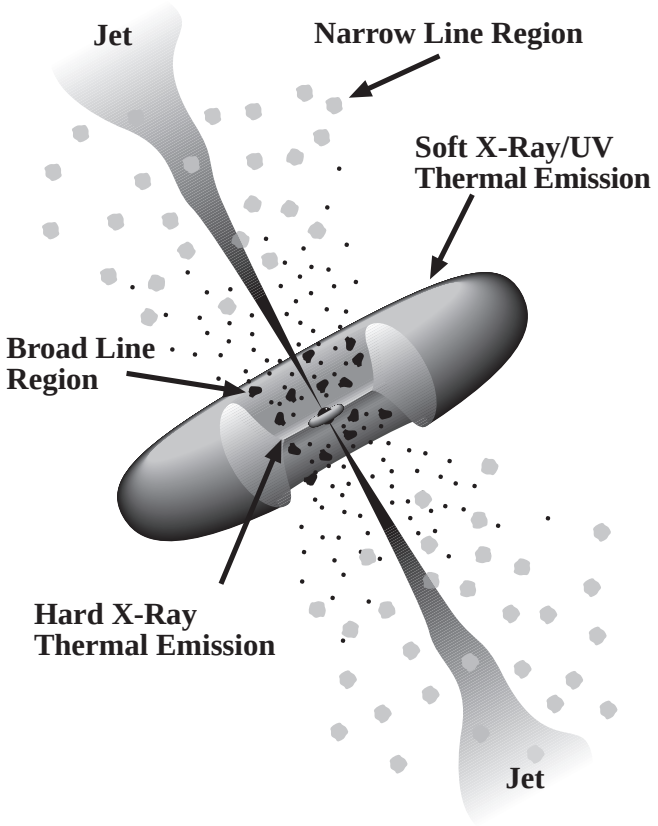

Figure 2.1: Schematic diagram of the basic AGN model (not to scale). A central supermassive *black hole* is surrounded by an *accretion disk* of heated plasma, emitting thermal UV and soft X-ray photons. Further to the outside is a *torus* of colder material (dust and molecular matter), obscuring the central parts of the AGN from direct view from shallow angles to the plane of rotation. The *broad-* and *narrow-line regions* contain clouds of gas, ionised by the radiation from accretion disk and jet, and moving at speeds higher (broad) or lower (narrow) than 2000 km s^{-1}. Bipolar, magnetically collimated *jets* of relativistically moving plasma are ejected perpendicularly to the accretion flow. Adopted from Urry & Padovani (1995).

For more information regarding the model, the reader is referred to Urry & Padovani (1995); a more physical treatment of the phenomena relevant for high-energy emission will be presented in Sec. 2.2.

Supermassive black hole

As motivated in the preceding, the AGN model is based on the existence of supermassive black holes (henceforth SBHs) with masses from typically $\sim 10^7$ up to $10^{10}\ M_\odot$ (see also Sec. 2.1.1). In order for a supermassive black hole to form and to stay active through the continuous accretion of matter, a surrounding reservoir of material must be available. This is the reason why indications for *active* supermassive black holes have so far only been found in the centres of galaxies, where the matter density is highest.[5] For a typical AGN with a black hole mass of $10^8\ M_\odot$, the radius of the event horizon is around $3 \cdot 10^{11}$ m.

Accretion disk and dust torus

Surrounding the central black hole is an accretion disk formed by matter spiralling into the black hole. The accretion disk is thought to draw its material from surrounding gas and dust, and occasional stars. The viscous and turbulent processes inside the disk heat the accreting gas to very high temperatures, giving rise to thermal UV and soft X-ray emission; the ultraviolet light emitted is thought to be responsible for the "blue bump" in the spectra of quasars. Via the same process, the rotating gas sheds angular momentum, allowing the innermost layers of the disk to fall towards the event horizon while new material enters the accretion disk from a thick co-rotating *dusty torus* (or similarly shaped axially symmetric distribution of dust and cold molecular matter), to which the thin accretion disk is joined at larger radii. It can be expected that the transition between ring and disk is not discrete, but rather gradual; as the temperature increases towards smaller radii along with the gravitational gradient, the dust evaporates, and the thickness of the accretion disk gradually decreases as the rate of inelastic collisions between particles increases. The existence of the torus has the important effect of obscuring the accretion disk and the regions immediately above and below it from an observer viewing the disk at shallow angles, but may also re-emit the absorbed radiation and hence give rise to emission lines. For our typical AGN of $10^8\ M_\odot$, the inner edge of the accretion disk is located at a radius of about $(1\text{-}30) \cdot 10^{12}$ m. At what exact distance from the black hole the dusty torus commences is very uncertain,

[5] There exists also no evidence for *inactive* (*i.e.* dark) SBHs being present anywhere else but in the centres of galaxies.

and probably depends on the level of activity of the AGN as well as parameters such as the magnetic field strength. For $M = 10^8\,M_\odot$ as assumed above, Urry & Padovani (1995) consider a radius in the order of 10^{15} m to be realistic.

Cloud regions

Space on either side of the accretion disk is filled with clouds of gas, which are trapped in the gravitational field of the black hole and rotate with it. These clouds are ionised by the radiation from the accretion disk, and give rise to strong line emission at optical and UV wavelengths. The widths of the observed lines are not due to thermal broadening, but Doppler broadening; one distinguishes between *narrow-line* and *broad-line* regions, depending on whether the velocity of the local gas clouds is smaller or larger than 2000 km s^{-1}, giving rise to narrower or broader emission lines, respectively. The broad-line regions are thought to be located very close to the black hole and to the inside of the dust torus – by which they are hidden from direct view at small angles to the plane of rotation – where the orbital velocity is highest. The strongest lines emitted are the Balmer series of hydrogen, and no forbidden lines are emitted, suggesting that the broad-line clouds are very high density (10^9 - 10^{13} ions/cm^3). Narrow lines are produced by the clouds further away from the black hole, where they are not obscured by the dusty torus and can be observed from any direction. Forbidden lines from metastable excited states are observed among the narrow lines, indicating a lower density. Apart from the gas clouds, a *corona* of hot electrons is also thought to be present in the zones immediately above and below the accretion disk, perhaps extending all the way to the narrow-line region. This corona may be responsible for scattering some of the broad-line emission into the line of sight for transverse viewing angles, at which the direct broad-line emission is obscured by the torus. The narrow-line region may extend from $\sim 10^{16}$ m to about 10^{18} m from the accretion disk.

Relativistic jets

Significant radio emission is detected from only about a fifth of all resolved AGN; these AGN are also called *radio-loud*, the rest are *radio-quiet*.[6] High-resolution images of radio galaxies show that the vast majority of radio emission is produced inside bipolar jets shooting out from the central black hole of the AGN perpendicularly to the accretion

[6] Adopting the most common definition, radio-loud AGN have radio ($\nu = 5$ GHz) to optical (B-band) flux ratios greater than 10. Under this definition, the fraction of radio-loud AGN is about 15% to 20%. However, due to the somewhat imprecise nature of the definition which is based purely on observation techniques, this fraction increases along with the optical flux, reaching 50% for optically very bright AGN (see Urry & Padovani, 1995, and references therein).

disk, and/or huge radio lobes at the ends of the jets. The jet matter is extracted from the accretion disk with a possible contribution from the electron corona. While the exact composition of the jets is still under debate (for details see Sec. 2.2.3), it is quite certain that they contain charged particles and magnetic fields. The collimation of the jets is likely to be caused by spiralling magnetic field lines. The jet plasma is moving at highly relativistic speeds, decelerating with increasing distance from the black hole. The charged particles inside the jets are thought to be further accelerated to extremely high energies by mechanisms such as magnetic shocks travelling down the jets, giving rise to synchrotron radiation as well as gamma rays. A detailed description of these processes will be given in the following section.

It is generally assumed that some form of relativistic jets are present in both radio-loud and radio-quiet AGN, but that the jet intensity is greatly reduced in the latter, presumably because of low accretion rates. Radio jets have been observed on scales from 10^{15} m to several times 10^{22} m (0.03 pc to $\sim$1.5 Mpc), which is about one order of magnitude larger than the largest galaxies (Urry & Padovani, 1995). The left picture in Fig. 2.2 shows an artist's conception of how a radio-loud AGN might look like from up close when the accretion disk with dust torus is viewed edge-on.

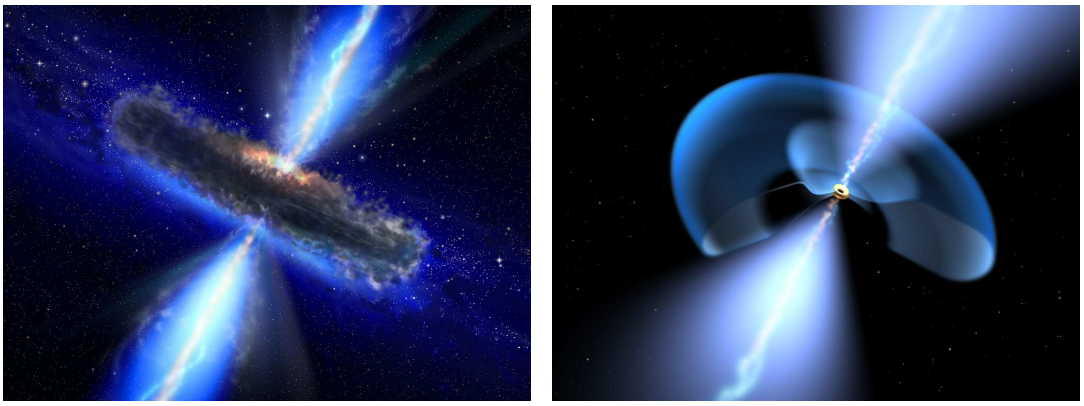

Figure 2.2: Artist's impression of a radio-loud Active Galactic Nucleus (AGN). *Left*: View from the side, showing the bipolar relativistic jets and the dusty torus which obscures view of the accretion disk. *Right*: Sketched view from a smaller angle to the jet, depicting the rotating accretion disk including the advection dominated bulge at the inner limit of the accretion disk (small yellow torus) and the large dust torus on the outside. The image scale is on the order of one light year. Credits to ESA/NASA (2006).

2.2 High-energy processes in AGN

In this section we shall take a closer look at those aspects of the AGN paradigm which are relevant for the acceleration of charged particles. Special emphasis is given to the relativistic plasma flow in the AGN jets, and models for the production of the non-thermal emission spectrum that is observed from these objects, which spans ~ 20 decades of energy from radio waves all the way to gamma rays.

2.2.1 Supermassive Black Holes

It is now believed that the centres of most, if not all galaxies harbour very massive black holes (see *e.g.* Kormendy & Gebhardt, 2001). The strongest evidence for the existence of SBHs, which boast masses between 10^6 and 10^{10} $M_\odot$, comes from observations of our own galaxy's centre. There, a very massive and compact gravitational attractor was discovered coincident with the position of the unusual radio source *Sagittarius A**. The rapid motions of several stars in the immediate vicinity of Sgr A*, which orbit around the invisible attractor with velocities as high as 1350 km s^{-1}, suggest that the compact object has a mass of between 2.6 and $3.7 \cdot 10^6\, M_\odot$. An upper limit on the size of the object can be obtained from the closest approach of the orbiting stars, yielding 45 AU (see Ghez, Salim *et al.*, 2005, for a recent measurement). The density resulting from the combination of these two numbers is too high even for a neutron star, hence the compact object must be a black hole. There is still some debate on how exactly supermassive black holes form in the centres of galaxies. A higher than average gas density and low galactic orbital velocities may give rise to an increased probability of star and gas mergers, and eventually the formation of a black hole, which will grow by accreting more matter. In the next section we will go into more detail as far as this accretion process is concerned. A useful review of SBHs can be found in Laor (1999).

In order to describe the effect of gravity on a particle in the vicinity of a black hole, Newtonian physics needs to be abandoned in favour of General Relativity, where gravitational forces arise from the local geometry of spacetime rather than from a force between point masses. Building on the introduction to black holes given in Sec. 2.1.1, no matter or photons can escape from within the event horizon of a black hole. The metric describing the curvature of spacetime around a *non-rotating* black hole, henceforth also referred to as *Schwarzschild black hole*, is the *Schwarzschild metric*

$$(\mathrm{d}s)^2 = \left(c\,\mathrm{d}t\sqrt{1-\frac{2GM}{rc^2}}\right)^2 - \left(\frac{\mathrm{d}r}{\sqrt{1-2GM/rc^2}}\right)^2 - (r\,\mathrm{d}\theta)^2 - (r\,\sin\theta\,\mathrm{d}\phi)^2 \quad (2.1)$$

(in polar coordinates with the black hole located at $r = 0$). Here G is the gravitational constant and M the black-hole mass. The radius of the event horizon of a non-rotating black hole is equal to the Schwarzschild radius $R_{\rm S}$, which can be obtained from Eq. 2.1 or deduced from purely Newtonian physics by requiring that the classical escape velocity equals the speed of light:

$$R_{\rm S} = \frac{2\,GM}{c^2} \tag{2.2}$$

Note that at $r = R_{\rm S}$, the time interval $\mathrm{d}t$ in Eq. 2.1 vanishes for all values of $\mathrm{d}r$, predicting that time stands still at the event horizon.[7] For a SBH with a mass of $X \cdot 10^8$ $\mathrm{M}_\odot$, the radius of the event horizon (and thus the effective "size" of the black hole) is about $X \cdot 300$ million kilometres. From Eq. 2.1 it follows directly that all radiation originating from a radius r_0 sufficiently close to the black hole is redshifted according to

$$\frac{\nu_{\rm obs}}{\nu_0} = \sqrt{1 - \frac{2\,G\,M}{r_0\,c^2}} \tag{2.3}$$

when observed at a large distance, where ν_0 is the emitted and $\nu_{\rm obs}$ the observed frequency. This relation is also called the *gravitational redshift*. It at first appears trivial, but is indeed quite remarkable that for non-relativistic motion, Eq. 2.1 implies an orbital frequency $\nu_{\rm orb}$ of a massive body around a Schwarzschild black hole that is identical to the prediction of Newtonian physics (*i.e.* Kepler's third law), namely

$$\nu_{\rm orb} = \frac{1}{2\pi}\sqrt{\frac{G\,M}{a^3}} \tag{2.4}$$

where a is the semi-major axis of the elliptical orbit.[8] Contrary to bodies in Newtonian physics, however, black holes cannot be orbited arbitrarily close. The metric gives rise to an *innermost stable orbit* below which radial perturbations disrupt the orbital motion of matter and cause it to fall onto the event horizon almost instantly. For circular orbits, the innermost stable orbit of a Schwarzschild black hole is $r_{\rm isco} = 3\,R_{\rm S}$ (see *e.g.* Bekenstein, 2005).

While all of the above properties pertain only to non-rotating black holes, real black holes resulting from stellar collapse or accretion of matter in the centres of galaxies are expected to have non-zero angular momentum. As already stated in Sec. 2.1.1, a rotating black hole with zero charge, or *Kerr black hole*, possesses an ellipsoidal event

[7] Equivalently, the gravitational redshift of a photon originating at $r = R_{\rm S}$ is infinite, which is why the event horizon of a Schwarzschild black hole is sometimes also called *surface of infinite redshift*.

[8] While this may come as no surprise for large values of a, it is not so self-evident for closer orbits where the curvature of spacetime is substantial.

horizon (Kerr, 1963). Unlike in the Schwarzschild case, the Kerr event horizon is *not* coincident with the surface of infinite redshift, which, except at the two poles, is located further outside. Furthermore, the infinite redshift surface is surrounded by an ergosphere where so-called *frame dragging* occurs, *i.e.* the rotation of three-dimensional space itself in such a way that all particles are forced to circulate in the angular directions, in the same orientation as the black hole spin. This is necessary in order to avoid $(\mathrm{d}s)^2 > 0$. The axial symmetry of the whole structure (in ϕ-direction) is of course retained.

The spacetime metric applying to Kerr black holes is somewhat more complex than that for Schwarzschild black holes and shall not be presented here; the interested reader is referred to Misner (1973). The Kerr metric contains a parameter that represents the angular momentum of the black hole per unit mass, $a \equiv L_{\mathrm{spin}}/M$; the product $ac^{-1} = L_{\mathrm{spin}}/Mc$ is also called *Kerr parameter* and carries the unit of length. The so-called *spin parameter* $a_\star \equiv ac/GM = L_{\mathrm{spin}}c/GM^2$ determines the geometry of the black hole: the larger $a_\star$, the more aspherical the metric. At the same time the metric restricts the allowed values of $a_\star$ to the range between 0 and 1, thus limiting the possible value of the angular momentum per square of the black hole mass; for $a_\star = a = 0$ (*i.e.* zero angular momentum), the Kerr metric simplifies to the Schwarzschild metric. The case $a_\star = 1$ represents a *maximally rotating black hole*, which has the angular momentum

$$L_{\mathrm{spin}}^{max} = \frac{G\,M^2}{c} \tag{2.5}$$

Angular momenta larger than L_{spin}^{max} (*i.e.* $a_\star > 1$) do not lead to a well-defined horizon radius. Just like for the Schwarzschild case, orbits close to the event horizon of a rotating black hole are unstable, however the radius of the innermost stable circular orbit decreases with increasing angular momentum from $3\,R_{\mathrm{S}}$ at $a_\star = 0$ (Schwarzschild limit) to $\approx 0.5\,R_{\mathrm{S}}$ at L_{spin}^{max} (*i.e.* for $a_\star = 1$). The latter is possible because the horizon distance also decreases with increasing L_{spin} (albeit the horizon shape becomes more oblate), allowing for orbits closer and closer to the event horizon.

2.2.2 Accretion disks

Apart from the much larger dimensions involved, the process of accretion via viscous accretion disks should be very similar in the supermassive black holes of AGN and the stellar mass black holes of X-ray binaries and microquasars. Provided there is a sufficient supply of matter in the vicinity of the black hole, which may include interstellar gas as well as matter from stars that are tidally disrupted in the strong gravitational field, the black hole will eventually consume surrounding matter and gradually become more

massive in the process. Indeed, discrete X-ray absorption lines around 0.6 keV (at higher or lower energy, depending on the AGN's flux state) provide evidence for low-density highly ionised gas surrounding the central engines of BL Lac objects (*e.g.* Sambruna *et al.*, 1997; Sambruna & Mushotzky, 1998). If accreting material can dissipate energy so that it is not pressure supported, it will settle into an accretion *disk*, which corresponds to the minimum energy configuration for a fixed angular momentum. The initial angular momentum will come from the rotation of the host galaxy, *i.e.* from infalling gas, dust and stars caught by the gravitational field of the black hole and forced into orbits around it.

Although the crucial role of accretion in powering AGN and other astrophysical objects was recognized early, several important aspects of the very nature of accretion disks are still puzzling. The most fundamental of these is the exact origin of the viscous stresses that transport angular momentum outward and allow matter to fall below the innermost stable orbit and through the event horizon. Ordinary molecular viscosity, for instance, is orders of magnitude too weak to play a significant role in accretion disks, as the densities are too low. Following the assumption that accretion disks are threaded by magnetic fields, Balbus & Hawley (1992) were the first to propose that the necessary viscosity could be generated by turbulence arising from these differentially rotating magnetic fields, causing so-called *magneto-rotational instabilities* (MRI). According to the theory of magneto-hydrodynamics (MHD), MRIs occur when a plasma disc rotates non-rigidly in a weak magnetic field. Since the rotational motion of the accretion disk matter around a black hole reaches relativistic speeds (in the inner parts of the disk, $v \gtrsim 0.1\,c$), this process is thought to be highly efficient. Detailed and very time consuming computer simulations, which have only recently become possible, are necessary to predict the magnitude and distribution of stresses due to MRI. While a qualitative agreement with observations has been established, a quantitative validation of the theory is still awaited as MHD simulations mature to the required level of detail.

In general, the total luminosity of an accretion disk L_{acc} can be expressed in terms of the accretion rate $\dot{M}$ via

$$L_{acc} = \epsilon\, \dot{M}\, c^2 \tag{2.6}$$

The proportionality constant ϵ is the (mass-to-radiation) energy conversion efficiency of the disk. In the conventional, slightly simplistic thin-disk model of accretion, all of the gravitational energy released through the accretion of matter is converted to radiation that escapes the disk. In this case, the total luminosity of the disk is thus just the accreted mass' difference in gravitational binding energy between the outermost and

innermost stable orbits of the disk, per unit time.[9] For a Schwarzschild black hole, this luminosity corresponds to an efficiency $\epsilon \approx 0.08$.

Assuming the accreting disk matter is a plasma of coupled electrons and ions, one can place an upper limit on the luminosity of the disk by balancing outward directed radiation pressure on the electrons due to Compton scattering in the Thomson regime with the gravitational force of the central mass M on the ions. Making the reasonable approximation that most of the ions are protons (*i.e.* hydrogen), one obtains the so-called *Eddington luminosity*

$$L_{\rm Edd} \;=\; \frac{4\pi\, c\, G\, M\, m_p}{\sigma_{\rm T}} \quad = 1.3 \cdot 10^{38}\, \frac{M}{M_\odot}\ \ {\rm erg\ \ s^{-1}} \tag{2.7}$$

The constant $\sigma_{\rm T}$ is the *Thomson cross section* for electrons (see Eq. 2.26). For accretion rates that lead to higher luminosities than $L_{\rm Edd}$, the disk would cease to be dynamically stable, causing super-Eddington regions to be "blown away". Note that the Eddington limit only applies to spherically symmetric radiation.

Even under the assumption that the angular momentum transport is driven by magnetic turbulence, there are several fundamentally different theories pertaining to the actual structure of accretion disks. Most of these models are so called *sub-Eddington* models where the emission is thought to originate from the disk itself, and whose luminosity and accretion rate are therefore constrained by the Eddington limit. The classical accretion disk model is that of Shakura & Sunyaev (1973), which assumes that the accretion disk is geometrically thin (*i.e.* at all radii, the vertical thickness is much smaller than the radius), but *optically thick*. The latter term means that the opacity of the disk is high enough so that it is held in thermal equilibrium by internal radiation. Such disks are cooled radiatively in an efficient manner, emitting blackbody spectra. Since the orbital velocity gradient increases with smaller radii, the viscous forces are highest at the inner edge of the accretion disk, as is the disk temperature. Therefore the total emitted spectrum will be a superposition of blackbody spectra of different temperatures. The efficiency of this type of disk is somewhere between 0.08 (for a non-rotating black hole; see above) and 0.42 (for a maximally rotating black hole).[10] The existence of a hot electron corona on either side of the accretion disk is supported by the observed disk emission from Seyfert galaxies, which appears to have two distinct components: a

[9] In this simplistic picture, once below the innermost orbit, the gas is assumed to decouple and fall onto the event horizon rapidly and under emission of only a negligible amount of radiation, taking the remaining angular momentum with it. In reality it may stay magnetically connected to the plasma in the accretion disk, and so increase the energy conversion efficiency ϵ further.

[10] AGN powered by thin-disk accretion are thus one of the most efficient energy sources of the universe. For comparison, in nuclear fusion it is only $\epsilon \approx 0.007$.

thermal component in the UV ("blue bump", assumed to come from the inner region of the disk itself) and a non-thermal, power law component in X-rays (UV photons which are inverse-Compton scattered to X-ray energies by hot electrons in the corona outside the disk; see *e.g.* Nayakshin, 1998).

In addition to the standard thin-disk model, geometrically thick (*i.e.* toroidal) disks are predicted to form under certain conditions. When the internal radiation pressure becomes significant enough to counteract self-gravity, the disk may "puff up" into a torus. This might happen at locations of the disk where the luminosity is high, such as at small radii close to the black hole. Another possible scenario for the formation of a thick torus is if the plasma density is so low that energy exchange between ions and electrons via collisions becomes inefficient. The electrons thus cool radiatively, while the ions remain at a high temperature. The resulting thermal pressure counteracts the disk's self-gravity and gives rise to a so-called "ion-supported torus" (Rees *et al.*, 1982).

Lately, another type of accretion disk model has gained popularity with many theorists, and that is the *Advection Dominated Accretion Flow* (short, ADAF) model. Related to the ion-torus idea outlined above, the ADAF principle was first proposed by Ichimaru (1977), but only received attention following the work of Narayan & Yi (1994). In contrast to the thin-disk accretion model, ADAFs involve an *optically thin*, but *geometrically thick* disk.[11] The optical thinness is the result of a low gas density, which allows radiation to escape; the toroidal geometry is caused by the same effects as in the "ion-torus" model outlined above. Just as radiatively inefficient as the former, the bulk of the energy is stored in the thermal motion of the ions rather than converted into radiation. This results in very high (nearly virial) disk temperatures on one hand, and unusually low luminosities on the other, as most of the energy is *advected* with the accretion flow rather than radiated away. Observationally, ADAF disks are interesting for several reasons. Firstly, their low luminosities may explain the extremely *underluminous* accretion disks in some sources, such as the one around the black hole in the centre of our own galaxy. Secondly, the expected slow rotational speeds, which are much less than Keplerian, prevent accreting objects from being spun up to unobserved high angular momenta. Finally, advective disks are susceptible to generating bipolar outflows, which may provide a natural source of matter for relativistic jets. Initially many experts doubted the stability of the proposed mechanism (see *e.g.* Nayakshin, 1998, and references therein), but since then MHD theory and numerical simulations have improved

[11] Optically thin disks seem to solve a number of problems encountered in optically thick accretion disk models, such as the fact that X-ray spectra from Seyfert galaxies are observed to be hard, but the temperature in optically thick disks falls off with the radius like $\sim 1/r^{3/4}$.

significantly. Today ADAF scenarios are widely accepted as possible mechanisms; *e.g.* Oda *et al.* (2007) found steady solutions for optically thin disk models threaded by toroidal magnetic fields, including ADAFs. ADAFs seem to be particularly suited to explain objects which have very low accretion rates, as the latter are more likely to result in optical thinness.

Finally, for completeness it should be mentioned that there are models of extremely high (super-Eddington) accretion rates. One way to circumvent the Eddington limit is by relaxing the usually imposed conditions of cylindrical symmetry and steady state. Other proposed means are those employed in the radiatively inefficient, advection cooled "Polish Doughnut" and "slim disk" models. Since these models are currently not very widely used, they will not be described here.

The discussion of the structure of accretion disks should be concluded with the remark that since current observations do not allow any one of the models to be ruled out with certainty, real accretion disks may be a hybrids between existing models. For instance, some authors assume that the outer parts of common accretion disks are thin and radiatively efficient, while the inner parts are toroidal and advection dominated. Furthermore, the dominating mechanism might change with time depending on the accretion rate; for instance, Oda *et al.* (2007) find models that include a transition between hot radiation dominated and hot magnetic pressure dominated disks depending on the density of the accreted material.

One fundamental question regarding the accretion of matter onto a central object remains, namely the one concerning angular momentum. As matter with non-zero angular momentum is accreted, the central object is naturally being "spun up". While this may continue up to very high rotational speeds, neither stars nor black holes can possess arbitrarily high angular momenta; the former will eventually be torn apart by centrifugal forces, while the angular momentum of the latter is limited by general relativity (see Sec. 2.2.1). In the case of rotating (Kerr) black holes, there must therefore be a way for the central object to shed angular momentum. There are various theories about how this may happen. One possible mechanism for extraction of angular momentum from a Kerr black hole via magnetic fields is the *Blandford-Znajek* process (Blandford & Znajek, 1977). When magnetic field lines anchored in the spinning accretion disk penetrate into the ergosphere of the black hole (*e.g.* along with matter that is being accreted), they are twisted and toroidal components are generated. The latter give rise to poloidal electric fields and, in the presence of ionised matter, currents. The black hole's rotational energy can thus be transported away from the black hole through Poynting flux. Through the latter, the Blandford-Znajek mechanism may also be responsible for

powering extragalactic jets. The caveat of this elegant prescription is that its efficiency is relatively low, so that it may contribute significantly only when the emission due to accretion itself is low, *i.e.* for low accretion rates. It is furthermore restricted to the regions immediately around the event horizon (Li, 2000). In general, the power that could be radiated via the Blandford-Znajek mechanism appears to be higher in the case of geometrically thick disks such as ion-tori or ADAFs than for thin disks.

2.2.3 Relativistic jets

When examining the non-thermal X-ray spectra of radio galaxies, it soon became apparent that most of them could not be explained only by emission from the accretion disk, with or without an electron corona. The double-lobed structure of the radio-emitting regions in images of radio galaxies hinted at jet-like outflows being present in these sources, and led Blandford & Rees (1978) to suggest that at all wavelengths, the radiation observed from so-called *blazars*, which are AGN whose jet axes are oriented at small angles to the line of sight, is dominated by non-thermal emission from relativistically moving astrophysical jets. Today, the existence of relativistic jets in AGN is thought to be a prerequisite for the acceleration of charged particles to energies high enough to give rise to the observed X-ray and gamma-ray emission, and in many cases there is ample evidence for their existence even when the jets themselves are not directly visible. Fig. 2.3 is a radio image of the nearby radio galaxy Cygnus A, in which the bipolar jets and the large lobes of radio emission at their ends can be seen particularly clearly.

Astrophysical jets are observed in direction of the rotation axis of accreting black holes (*i.e.* perpendicular to the accretion disk). They are generally assumed to be bipolar; cases where only one jet is observed can usually be explained by differences in luminosity, most often caused by relativistic beaming effects due to a non-perpendicular viewing angle. The emission spectra of blazars, which are dominated by relativistic beaming effects due to very small jet viewing angles, are of particular importance to this work; a more detailed treatment of the relevant physical processes within their jets will follow in Sec. 2.3.

Jet formation

Neither the exact composition of astrophysical jets nor the mechanism through which they are driven and collimated is currently known with certainty. There are a number of models for the mechanism responsible for launching the jet and keeping it collimated up to distances of a Mpc from its origin. It is believed that the force on the charged particles

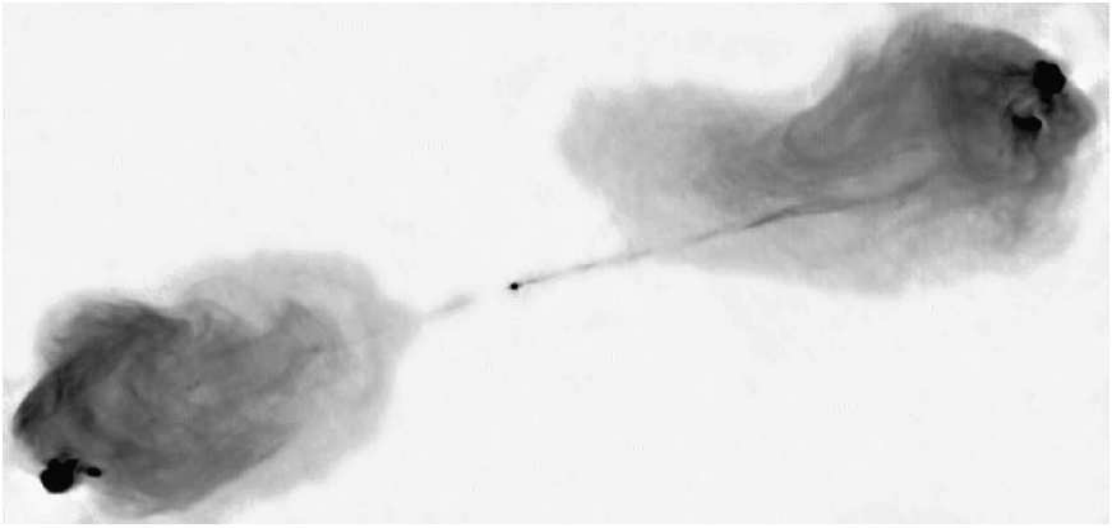

Figure 2.3: The FR II radio galaxy *Cygnus A* as seen at radio frequencies (5 GHz) by the *Very Large Array* (VLA) of radio telescopes in New Mexico, USA (Carilli & Barthel, 1996). The field of view of this negative image is 2.3' × 1.3', the resolution 0.5 arcsec. The image shows that the highest radio intensity is emitted by the big radio lobes at the ends of the relativistic jets, in particular in several *hot spots*. Radio emission from the jets, which are also visible, is comparably faint. The jet axis is not exactly perpendicular to the line of sight, but slightly inclined towards the observer, hence the approaching jet (to the right) appears to be brighter than the receding counterjet.

is transmitted by ordered magnetic fields that thread the accretion disk, in which case jet physics are governed by magnetohydrodynamics. It is most intuitive to assume that the jet matter is made up of charged particles from the accretion disk, *i.e.* an electron-ion plasma, which are magnetically accelerated and collimated in the direction perpendicular to the plane of rotation of the black hole. This concept is also called *disk wind* (see *e.g.* Blandford & Payne, 1982). Another possibility is that the jets are dominated by the Poynting flux from vacuum discharge, according to the Blandford-Znajek mechanism (see previous section and Blandford & Levinson, 1995). In this case, jets would consist predominantly of electrons and positrons, and few to no hadrons. In both cases, the jet is thought to carry along a "frozen in" magnetic field. Depending on the model, the kinetic energy of the ejected jet plasma may come from the angular momentum of the accretion disk or that of the black hole itself (see also previous section), or from the released gravitational energy of the accretion disk.

From the observation of superluminal motion (see later in this section) and fits of gamma-ray emission models to the broad-band *spectral energy distributions* (SEDs) of blazars, the bulk Lorentz factors Γ of the plasma within relativistic jets of AGN appear

to be in the range of about 8 - 40. Observational evidence suggests that the acceleration of the jet matter is not quasi-instantaneous, but occurs gradually over a distance of typically several parsecs (Sikora *et al.*, 2005), starting at the launching point of the jet observed to be as close as less than a parsec from the central black hole. In addition to this relativistic bulk motion, the jet particles are also subject to thermal motion within the rest frame of the emitted plasma. Like their formation, the collimation of the jets is likely achieved magnetically. The rapid rotation of the disk causes the magnetic field lines to become twisted even large distances from the ergosphere of the Kerr black hole. As charged particles spiral around the field lines, the toroidal component of the spiralling magnetic field serves to collimate the particle stream as it propagates away from the black hole. The effectiveness of the collimation process depends on the relativistic speed of the jet, and is approximately $\theta \sim 1/\Gamma$, where θ is the jet cone opening angle.

Acceleration mechanisms

As we will see in Sec. 2.3, all current models of gamma-ray production in AGN require the acceleration of charged particles to very high energies (for electrons, $\gtrsim 10^9$ - 10^{13} eV). The inferred bulk Lorentz factors of AGN jets ($\sim$ 10 - 40) correspond to energies of 5 - 20 MeV for electrons and 9 - 38 GeV for protons, which is several orders of magnitude too low to produce the observed gamma-ray emission in the HE to VHE energy band. It follows that there must be another mechanism at work *within* the jets which accelerates charged particles further to ultra-high (TeV and above) energies.

The question of how exactly particle acceleration takes place within astrophysical jets, *i.e.* what physical processes are involved and how they contribute, cannot be answered with certainty at present. While it is easiest to assume that moving magnetic fields provide the necessary energy boost for the charged particles, it is still not clear how exactly this happens. The most popular acceleration mechanism invoked to explain the inferred energy distribution of the particles is the magnetically driven *first-order Fermi acceleration*, also referred to as *shock acceleration* (Blandford & Eichler, 1987). According to MHD theory, shock acceleration occurs when a faster moving plasma encounters a slower moving plasma at a relativistic speed greater than the speed of sound within the slower plasma. In the interaction region of the two fluids a *magnetic shock* develops, within which charged particles that are trapped in it may be repeatedly magnetically reflected between the leading and trailing boundaries of the shock. Since in the rest frame of the shock, both boundary surfaces approach each other, the particles will *gain* energy each time one of the shock fronts is traversed. At one point the particles will either escape the shock or lose energy by means of inelastic collisions. With increasing energy

due to acceleration, the gyroradius of the accelerated particles increases, resulting in a higher probability of escaping the shock. This mechanism naturally produces a power law energy spectrum of particles (Eq. 2.31), with the value of the exponent depending on the shock compression ratio. For reference in the following sections, we shall denote the spectral index α of the accelerated particles by p. In the non-relativistic case, the power law index p is thought to be around 2.0, however in the relativistic limit slightly steeper spectra are assumed, with p in the range of 2.2 - 2.3 (Gallant, 2002). It can be shown that when parameters such as pitch angle are modelled in detail, the range of allowed indices can be extended to $1.5 \lesssim p \lesssim 3.0$ (Blasi & Vietri, 2005). Among others, MHD shocks are thought to be produced within astrophysical jets by plasma "blobs" that were ejected with different velocities (internal shocks), or at the boundary surface where the jet is driven into the surrounding medium (external shocks). Both electrons and ions (protons) can likewise be accelerated at the shock fronts.

So-called *second-order Fermi acceleration* (Fermi, 1949) may also contribute to the acceleration, although it is thought to be less efficient than shock acceleration. In this case, acceleration is achieved by charged particles being scattered elastically on moving inhomogeneities or bends in the local magnetic field lines. Depending on whether they are reflected by approaching or receding magnetic fields, they may gain or lose energy. However, since for relativistically moving particles the former case occurs more frequently than the latter, the net result is a gradual energy gain. This stochastic process may be able to accelerate charged particles in turbulent magnetic fields, or during reconnection of field lines.

Jet structure

The morphology of many AGN jets viewed at large angles (*i.e.* in radio galaxies) comes to light owing to their significant radio emission, whose large degree of linear polarisation suggests that it is synchrotron radiation from charged particles gyrating within the magnetic fields in the jet. The structure of the jets is anything but uniform. Bright spots, "blobs" or "knots" can be seen in the radio images of almost all jets. Many jets (particularly the weaker jets of FR I radio galaxies, see Sec. 2.4.1) start out straight but become curved some distance away from the launching point. The bright spots in the jet are most likely the sites of internal magnetic shocks; in luminous radio galaxies, they stand out through the fact that – according to polarisation measurements – the local magnetic field is perpendicular to the jet axis in their interior, whereas the large-scale magnetic fields on the jet surface and in the less active part of the jet appear to be predominantly parallel to the jet (Kembhavi & Narlikar, 1999; Perlman *et al.*, 1999).

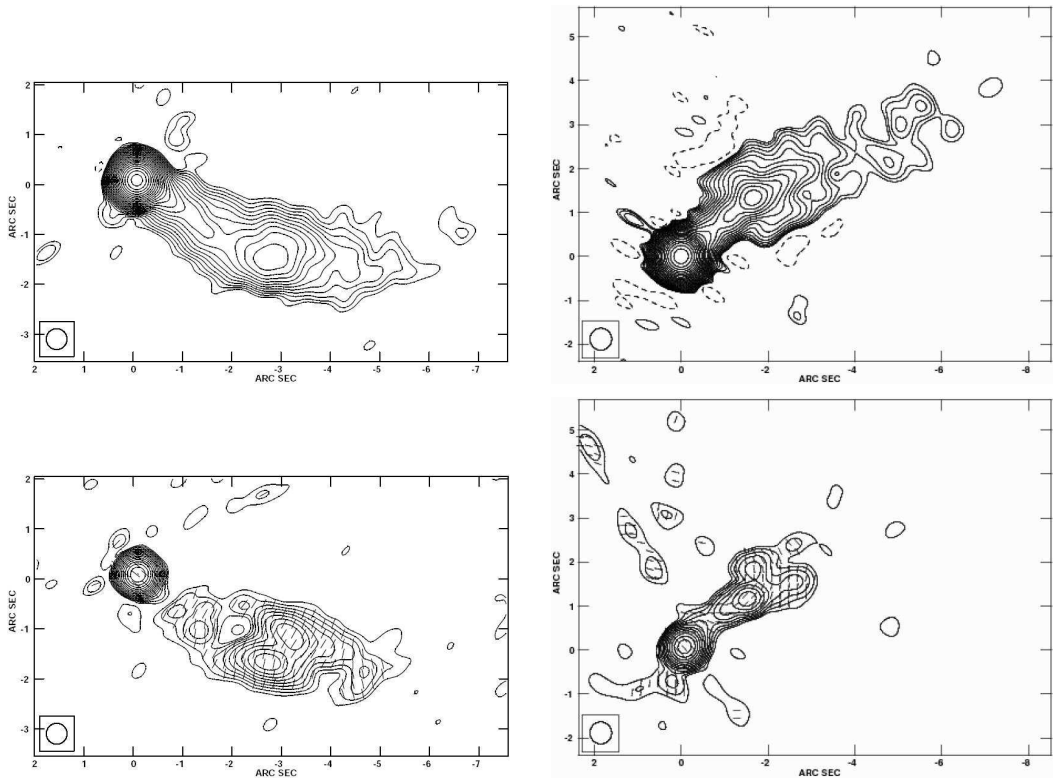

Figure 2.4: Radio images of the central regions of the jets of the two FR I radio galaxies 3C 371 (at 5 GHz, *left*) and PKS 2201+044 (at 8.5 GHz, *right*), from observations with the VLA. *Top*: Total intensity maps. *Bottom*: Polarisation maps showing the observed electric field. The restoring beam of 0.42 arcsec is shown in the lower left corners. From Sambruna *et al.* (2007).

Alternatively, magnetic outbursts or reconnection of field lines may result in the ejection of discrete blobs of material, in which case the bright knots would correspond to areas of enhanced gas density. A significant curvature of the outer, slower moving jet parts can occur as a result of the galaxy moving through the local cluster, where the jets are swept back and distorted through ram pressure of the intra-cluster gas. Radio images including polarisation maps of the central parts of the FR I radio galaxies 3C 371 and PKS 2201+044 are shown in Fig. 2.4. Only the bright knots of these jets show up in the polarisation maps, with polarisation predominantly perpendicular to the jet axis.

Apart from being decelerated by the surrounding medium – from highly relativistic speeds close to the black hole to rather low velocities near the termination points – astrophysical jets are also thought to have a transverse velocity gradient, with the velocity being highest in the centre and falling off towards the outside of the jet. The large radio lobes often seen in images of radio-loud AGN with luminous jets (see Fig. 2.3)

are caused by the jet particles being slowed down by the gas inside of and surrounding the galaxy, causing a turbulent back-bouncing and so-called *hot spots* where the inner high-velocity components of the jet terminate in high-density regions of the surrounding medium. The energy stored in the radio lobes is immense; each lobe of the jets of Cygnus A (see Fig. 2.3) is estimated to contain 10^{60} - 10^{61} erg of energy, which is thought to be equally shared between the particles' kinetic energy and magnetic energy (Carrol & Ostlie, 1996). This is equivalent to the energy released by $\sim 10^7$ supernovae. Fig. 2.4 shows an X-ray image of the jet of prominent nearby radio galaxy M 87, a VHE gamma-ray emitter. A radio image is overlaid, demonstrating the spatial coincidence of the X-ray and radio emission of the various knots.

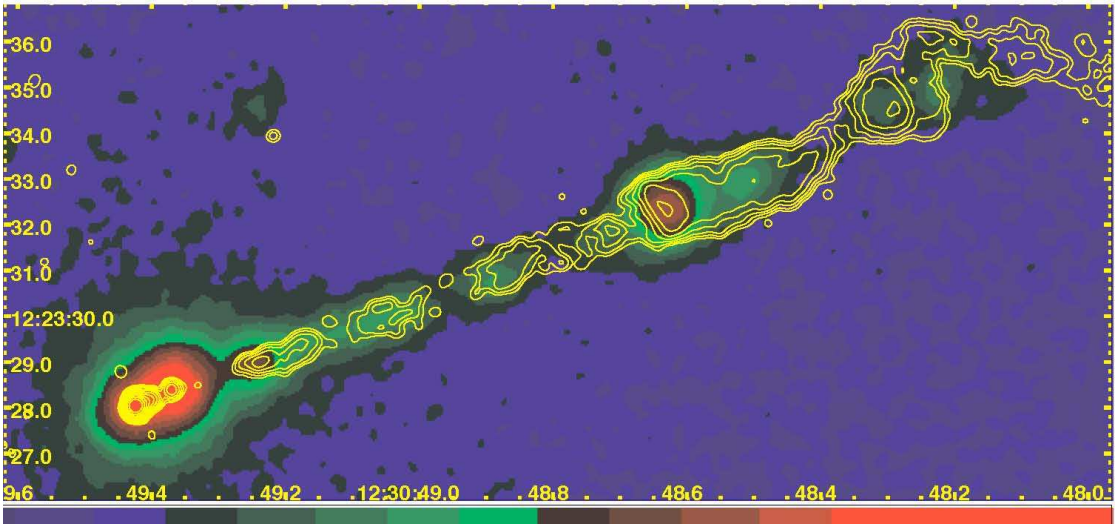

Figure 2.5: X-ray image of the jet of the radio galaxy M 87 as seen by Chandra (colour contours, for relative luminosities see the scale at the bottom of the image). The overlaid radio contour lines are from a VLA observation at 8 GHz, with contours increasing by factors of two. The jet is approaching us at an angle of about 30°. The feature to the lower left is the nucleus, followed by the knots HST-1, D, E, F and A. HST-1 and A are the brightest in X-rays, second only to the area surrounding the nucleus. From www-Chandra (2008).

The bulk Lorentz factor decreases as the emission regions travel down the jet, owing to ambient gas and photon fields decelerating the ejecta. Acceleration of charged particles to the highest energies may therefore be expected to occur at the base of the jet, not far from the Schwarzschild radius of the black hole itself. This however is also the location where the density of ambient background photons, originating directly from the accretion disk or rescattered from broad-line clouds or dusty torus, is highest (Spada *et*

al., 2001). In the far upstream part of the jet, the efficiency of the acceleration process is hence reduced, and the spatial origin of the highest energy radiation within the jet will be dependent on factors such black hole mass, accretion disk luminosity and accretion rate.

Superluminal motion

One unique indicator for the presence of relativistic jets in an object is the observation of so-called *superluminal motion* of plasma blobs, as seen at radio or visual wavelengths. Despite its name, this effect does of course not violate special relativity, nor do relativistic kinematics need to be invoked for its explanation. It can be observed whenever the velocities of the blobs exceed a certain minimum velocity which depends on the angle θ between the blob's direction of motion (*i.e.* generally the direction of the jet) and the line of sight. If $v = \beta c$ is the velocity of the blob in the local rest frame, it can be shown in a straightforward manner using only geometry that the *apparent* (*i.e.* observed) transverse velocity $\beta_{\rm app} = v_{\rm app}/c$ of the blob is

$$\beta_{\rm app} \;=\; \frac{\beta\,\sin\theta}{1\;-\;\beta\,\cos\theta} \tag{2.8}$$

Superluminal motion can therefore be observed whenever the boundary condition

$$\beta \;>\; \frac{1}{\sin\theta\;+\;\cos\theta} \tag{2.9}$$

is fulfilled. The absolute minimum bulk velocity an emission region can have for superluminal motion to be observed is thus $\beta_{\rm min} = 1/\sqrt{2} \simeq 0.71$, under the condition that the viewing angle is $\theta_{\rm min} = 45^\circ$.

Beaming effects

Due to the relativistic motion of the jet plasma, observable quantities such as time and energy will change depending on whether they are measured in the emission or the observer's frame of reference. Transformations from the rest frame of the jet, which moves relativistically with a bulk Lorentz factor Γ, to the observer's frame or the local rest frame involve the *relativistic Doppler factor* δ such that

$$\Delta t_{\rm obs} \;=\; \frac{1}{\delta}\,\Delta t \qquad \nu_{\rm obs} \;=\; \delta\,\nu \tag{2.10}$$

when Δt (ν) is the proper time (frequency) of a photon in the frame of emission and $\Delta t_{\rm obs}$ ($\nu_{\rm obs}$) its proper time (frequency) as determined by the observer at rest. The relativistic

Doppler factor is defined as (see *e.g.* Carrol & Ostlie, 1996, for the straightforward derivation)

$$\delta = \frac{1}{\Gamma\,(1 - \beta\cos\theta)} \tag{2.11}$$

where $\Gamma = (1-\beta^2)^{-1/2}$ is the bulk Lorentz factor of the jet, $\beta = v_{\rm jet}/c$ the velocity of the emission region in the jet, and θ the angle between the jet axis and the line of sight. Note that due to time dilation, even transverse motion is affected by the relativistic Doppler effect, such as in the case of radio jets perpendicular to the line of sight. In this case, $\delta = 1/\Gamma \ll 1$. For blazars on the other hand, which have $\theta \lesssim 10^\circ$, the particles emitted within the approaching jet arrive with an energy $h\nu$ that is Doppler-boosted to energies an order of magnitude higher. This becomes clear when plotting Eq. 2.11 for several Lorentz factors, as it is done in Fig. 2.6. From fits using leptonic gamma-ray emission models (see Sec. 2.3.1), a sample of well observed blazars appear to have bulk Lorentz factors of $\Gamma = 10$ - 20 and typical Doppler factors of $\delta = 15$ - 35 (Ghisellini *et al.*, 1998; Tavecchio, Maraschi & Ghisellini, 1998).

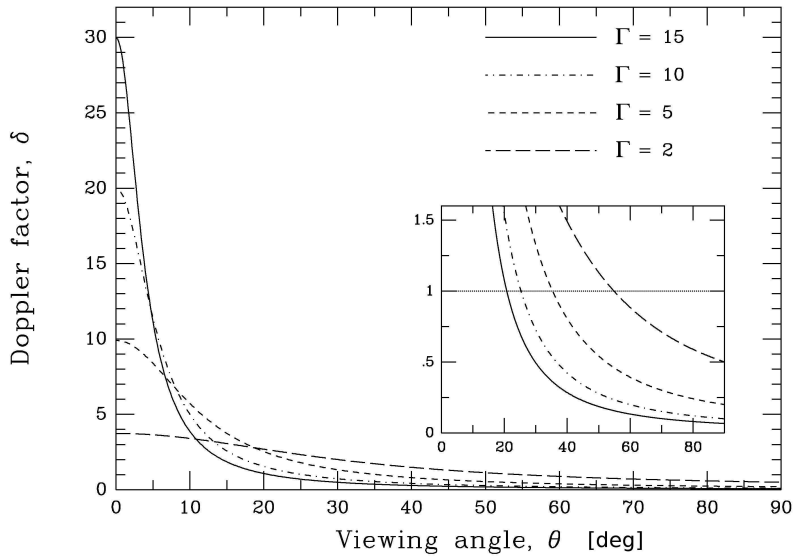

Figure 2.6: Relativistic Doppler factor δ as a function of viewing angle θ, for various different Lorentz factors Γ. The inset shows the angles for which $\delta = 1$ at an expanded scale. Adapted from Urry & Padovani (1995).

The particle energy is not the only observable that is Doppler-boosted. Resulting from the relativistic *headlight effect*, any initially isotropic emission is strongly beamed in direction of the jet motion as the solid angle element $\mathrm{d}\Omega$ transforms according to $\mathrm{d}\Omega_{\rm obs} = \mathrm{d}\Omega/\delta^2$. Another factor of δ applies due to the relativistic compression of the

time between the emission of individual photons (Eq. 2.10). The *observed* particle flux $\Phi_{\rm obs}$ is therefore

$$\Phi_{\rm obs}(\delta\, E) \,=\, \delta^3\, \Phi(E) \quad \Rightarrow \quad \Phi_{\rm obs}(E) \,=\, \delta^{3+\alpha}\, \Phi(E) \tag{2.12}$$

where Φ is the emitted flux measured in the moving frame of the jet, E the emitted particle energy, and $\delta\, E$ the measured (blue-shifted) energy. The right-hand identity follows if the emitted particle spectrum is a power law with spectral index α (Eq. 2.31). While above equation applies to continuous jets, the exponent $(3 + \alpha)$ is replaced by $(2 + \alpha)$ in the case of discrete emission regions (Blandford & Königl, 1979). From Eq. 2.12 it follows that for blazars, which appear to have typical Doppler factors of 20 and emit spectra with $\alpha = 1\text{-}2$, the emitted fluxes are amplified by five to more than six orders of magnitude, giving rise to apparent luminosities which are higher than the actual luminosities by the same amount. In these types of sources therefore, the beamed emission from the jet by far outshines the emission from all other parts of the AGN, including the bright accretion disk.

One very characteristic feature of blazars is the fact that their non-thermal emission is highly variable across the whole spectrum. From the time scale of the variability, determined by the rise or decay time Δt of observed flares (temporary, sometimes dramatic flux increases), an upper limit on the size of the emission region R can be inferred in a straightforward manner by arguing that the light-crossing time can be no larger than the variability time. This condition translates to

$$R \,\lesssim\, c\, \Delta t\, \frac{\delta}{1+z} \tag{2.13}$$

where z is the redshift of the source. For two blazars, gamma-ray flares with rise times as short as 3 minutes have recently been observed (PKS 2155-304 and Mrk 501; Aharonian *et al.*, 2007a; Wagner *et al.*, 2007), suggesting that $R/\delta < 5 \cdot 10^{12}$ cm. This condition requires either extremely compact emission regions of only a few Schwarzschild radii, or very large Doppler factors ($\delta \gtrsim 100$). Not only blazars exhibit short-term variability: variable VHE gamma-ray emission – uncorrelated with the fluxes at all other wavelengths – with $\Delta t \approx 2$ days was detected from the nearby ($z = 0.0043$, corresponding to $\sim$16 Mpc) FR I radio galaxy M 87 (Aharonian *et al.*, 2006d). From Eq. 2.13 is follows that $R \lesssim 5 \cdot 10^{15}\, \delta$ cm, which is about $5\, \delta\, R_{\rm S}$. The large-scale jets in this AGN are thought to have an angle of $\sim 30^\circ$ to the line of sight, suggesting a Doppler factor of less than one. If the emission region is in the immediate vicinity of the black hole, however, smaller angles θ and hence larger Doppler factors are considered possible. Either one of

these explanations challenge our current understanding of relativistic jets, and it is clear that more multiwavelength data and more detailed models are needed to answer these and the other remaining questions.

2.2.4 Synchrotron radiation

The observation that AGN jets emit highly polarised radiation supports the idea of large-scale magnetic fields being present within them. In this case, the observed emission is dominated by *synchrotron radiation*[12] emitted by very energetic charged particles that are accelerated in the magnetic fields.

A relativistic particle of mass m, charge q and Lorentz factor $\gamma = (1 - \beta^2)^{-1/2}$ travelling through a magnetic field of strength B will radiate with the frequency

$$\nu_{\rm s} = \frac{q\,B}{2\pi\,\gamma\,m\,c} \tag{2.14}$$

which can be derived readily by equating (relativistic) Lorentz and centrifugal forces. At relativistic energies, the emitted dipole radiation is strongly beamed in the forward direction, into a cone of half-angle $\theta \approx mc^2/E = \gamma^{-1}$. The frequency $\nu_{\rm s}$ is called the *synchrotron frequency*. For a uniform magnetic field, the particle trajectory will be that of a helix with radius

$$R_{\rm L} = \frac{\gamma\,\beta\,m\,c^2}{q\,B} \tag{2.15}$$

where $R_{\rm L}$ is called the *Larmor radius*, or *gyroradius*, of the particle. Note that the synchrotron frequency is independent of the pitch angle α between particle trajectory and magnetic field. As the particle energy increases, however, the emitted synchrotron photons will no longer be of one discrete frequency, but relativistic effects result in the emission of a distribution of frequencies. The frequency at which this distribution peaks is given by the spectral *cutoff* frequency

$$\nu_{\rm c} = \frac{3}{2}\,\gamma^3\,\sin\alpha\;\nu_{\rm s} = \frac{3}{4\pi}\,\frac{q\,B\,\gamma^2\,\sin\alpha}{m\,c} \tag{2.16}$$

The fact that the square of the Lorentz factor γ enters in the numerator is important, as this results in particles of higher energy emitting higher frequency synchrotron radiation. As we can see, the cutoff frequency increases with pitch angle as α is approaching 90°. The intensity P of the radiation (*i.e.* the emitted energy in the form of synchrotron photons per unit time) increases with the square of the perpendicular component of the

[12] In the non-relativistic case, this radiation is also called *magnetobremsstrahlung* or *cyclotron radiation*.

magnetic field (see *e.g.* Giacconi & Gursky, 1974):

$$P = \frac{2}{3}\frac{q^4}{m^2c^3}\gamma^2\beta^2 B^2 \sin^2\alpha \tag{2.17}$$

For a given velocity β, the emitted power is inversely proportional to m^2, hence it is evident that synchrotron radiation from electrons is far more important than from protons, which radiate efficiently only at much lower frequencies (see Eqs. 2.14 and 2.16). The synchrotron emission observed from AGN jets is thought to be exclusively due to electrons, and, depending on the actual jet composition (see Sec. 2.2.3), positrons.

Inside a relativistically moving jet, there will be an ensemble of electrons/positrons with randomised pitch angles radiating synchrotron photons. If one assumes that the injected charged particle spectrum is a power law with spectral index p as expected in the case of Fermi shock acceleration (see Sec. 2.2.3; note that the sign convention we shall adopt for all spectral indices α is the one given by Eq. 2.31), and averages over pitch angle, it can be shown that the superposition of the synchrotron photons also give rise to a power law emission. The spectral index of the synchrotron emission is then $s = (p-1)/2$ (see for instance Landau & Lifshitz, 1971), *i.e.*

$$\frac{\mathrm{d}N}{\mathrm{d}\nu}(\nu) \propto \nu^{-s} = \nu^{-(p-1)/2} \tag{2.18}$$

An accelerated electron spectrum of spectral index in the range $p = 1.5 \ldots 2.5$ will thus emit synchrotron radiation with a spectral index of $s \simeq 0.25 \ldots 0.75$. Note that these spectral indices apply to particle fluxes. Often the synchrotron radiation component is quantified in terms of *energy flux*, in which case the spectral indices are lower by one power (compare Eq. 2.31).

At sufficiently low energies, the plasma of charged particles becomes partially opaque to its own synchrotron radiation, and synchrotron emission is inhibited by *synchrotron self-absorption*. At the energy where synchrotron self-absorption becomes relevant, the emission spectrum turns over from the power law ($\propto \nu^{-s}$) above the break frequency to a spectrum varying as $\mathrm{d}N/\mathrm{d}\nu \propto \nu^{5/2}$ below it, causing a broad peak in the combined synchrotron spectrum of the ensemble.

One very important property of synchrotron radiation is the fact that it is polarised (it is in fact dipole radiation) with the electric field vector pointing in the direction of deflection of a negatively charged particle, and perpendicular to the primary magnetic field.

2.2.5 Inverse Compton Scattering

In the process known as *Inverse Compton scattering* (IC), a relativistic charged particle – most commonly an electron or positron, but the formalism also applies to other (heavier) particles – interacts electromagnetically with a low-energy photon ($\gamma_{\rm le}$), upon which a large fraction of its energy and momentum are transferred to the photon, boosting it to higher energies ($\gamma_{\rm he}$). The prefix "inverse" refers to the fact that in the chosen frame of reference (generally the observer frame), the charged particle transfers energy to the photon, instead of vice-versa as in ordinary Compton scattering. Thus the only difference between the two is the choice of reference frame. The Compton/IC process is most efficient for relativistic electrons/positrons:

$$e^{\pm}_{\rm he} + \gamma_{\rm le} \longrightarrow e^{\pm}_{\rm le} + \gamma_{\rm he}$$

The (unpolarised) differential cross-section of the Compton scattering process including relativistic effects is described by the *Klein-Nishina formula*

$$\frac{{\rm d}\sigma}{{\rm d}\Omega} = 0.5\, r_e^2 \left[P(E_\gamma,\theta) - P(E_\gamma,\theta)^2 \sin^2\theta + P(E_\gamma,\theta)^3\right] \tag{2.19}$$

where E_γ is the incident photon energy, θ the scattering angle,

$$r_{\rm e} = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{m_e c^2} \simeq 2.81794 \cdot 10^{-15}\,{\rm m} \tag{2.20}$$

the classical electron radius (m_e is the electron rest mass, e the electron charge and ε_0 the permittivity of free space), and

$$P(E_\gamma,\theta) \equiv \frac{E'_\gamma}{E_\gamma} = \frac{1}{1+\frac{E_\gamma}{m_e c^2}(1-\cos\theta)} \tag{2.21}$$

the ratio of the photon energy after and before the collision. Note that Eq. 2.19 is the special case for electrons/positrons. The generalised form of the Klein-Nishina equation, applicable to any charged particle, is

$$\frac{{\rm d}\sigma}{{\rm d}\Omega} = 0.5\, r_{m,q}^2 \left[P_m(E_\gamma,\theta) - P_m(E_\gamma,\theta)^2 \sin^2\theta + P_m(E_\gamma,\theta)^3\right] \tag{2.22}$$

Here $r_{m,q}$ is the "classical radius" of a particle with mass m and charge q,

$$r_{\rm m,q} = \frac{1}{4\pi\varepsilon_0}\frac{q^2}{mc^2} \tag{2.23}$$

and $P_m(E_\gamma, \theta)$ the ratio of the photon energies as in Eq. 2.21 but with m substituted for the electron mass m_e. In the low-energy limit ($E_\gamma \ll mc^2$), it follows that $\frac{E_\gamma}{mc^2} \to 0$ and the collision becomes elastic, yielding the cross-section formula for classical *Thomson scattering*,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; 0.5\, r_{m,q}^2 (1 + \cos^2\theta) \tag{2.24}$$

The Thomson limit corresponds to the case where the recoil of the charged particle is neglected. For electrons/positrons, the Thomson cross-section is often written in the form

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \;=\; \frac{3}{16\,\pi}\,\sigma_\mathrm{T}\,(1 + \cos^2\theta) \tag{2.25}$$

where the constant σ_T is the *Thomson cross-section for electrons*,

$$\sigma_\mathrm{T} \;\equiv\; \frac{8\pi}{3}\, r_e^2 \;\simeq\; 6.6525 \cdot 10^{-27}\,\mathrm{m}^2 \tag{2.26}$$

and r_e^2 the corresponding *differential* Thomson cross-section.

Comparing Equations 2.19 and 2.22 for electrons and protons, the ratio of the respective cross-sections is

$$\frac{(\mathrm{d}\sigma/\mathrm{d}\Omega)_e}{(\mathrm{d}\sigma/\mathrm{d}\Omega)_p} \;\simeq\; \left(\frac{m_p}{m_e}\right)^2 \tag{2.27}$$

confirming that the scattering amplitude is more than six orders of magnitude higher for electrons than for protons. IC-scattering is therefore an efficient energy loss mechanism for highly relativistic electrons, but not so for hadrons. This fact is of some importance in the context of gamma-ray production in relativistic jets, as will be explained in Sec. 2.3.2.

The kinematic region where the errors resulting from applying the Thomson approximation become significant ($\gtrsim 5\%$) is referred to as the *Klein-Nishina regime*. The distinction between Thomson regime and the highly energetic Klein-Nishina regime is of some importance in high-energy astrophysics. In the Thomson regime, the cross-section remains constant ($\sim \sigma_\mathrm{T}$) with energy; as a result, photons scatter very efficiently off of charged particles, in scattering amplitude as well as energy (*i.e.* quasi-elastically). With increasing energy the Klein-Nishina regime is entered, and the cross-section (Eq. 2.19) decreases rather rapidly along with the energy of the scattered photon, meaning the scattering becomes more and more inefficient. For electrons, Klein-Nishina effects become substantial already at photon energies as low as ∼10 keV for an electron at rest, corresponding to collisions between TeV-electrons and mid-infrared photons.

The final photon energy $E_{\gamma,\mathrm{he}}$, *i.e.* the energy of the up-scattered photon, can be most easily estimated by examining the scattering process in the rest frame of the charged particle, because then Equations 2.19 to 2.24 can be applied directly. Assuming the

incoming charged particle has an energy $E_{q,\mathrm{he}} = \gamma mc^2$, where $\gamma = (1-\beta^2)^{-1/2}$ is the Lorentz factor with $\beta = v/c$, the energy $E_{\gamma,\mathrm{le}}$ of the initial photon is transformed from the observer frame to the (primed) rest frame of the charged particle according to

$$E'_{\gamma,\mathrm{le}} = \gamma\,(1-\beta\cos\phi)\,E_{\gamma,\mathrm{le}} \tag{2.28}$$

where ϕ is the angle between the momenta of the charged particle and photon before the collision in the un-primed (observer) frame. In the rest frame of the charged particle, we assume the collision to be approximately elastic and let $E'_{\gamma,\mathrm{he}} = E'_{\gamma,\mathrm{le}}$ in accordance with Eq. 2.24. Note that for ratios of E_γ/mc^2 above $\sim 3\%$, the error incurred by making this assumption becomes significant as the Klein-Nishina regime is entered and the photon will be scattered with slightly reduced energy according to Eq. 2.21. For simplicity we ignore such corrections here, keeping in mind that for very high energies our result will be overestimating the correct value by some fraction of it. We obtain the final energy $E_{\gamma,\mathrm{he}}$ of the photon in the observed frame by applying the inverse Lorentz transformation (and also using the fact that $\beta \simeq 1$ to a very good approximation, in particular for electrons):

$$E_{\gamma,\mathrm{he}} \simeq 2\gamma^2\,(1-\cos\phi)\,E_{\gamma,\mathrm{le}} \tag{2.29}$$

This means that during IC-scattering in the Thomson regime, the photon's energy is boosted by the square of the charged particle's Lorentz factor. As the energy of the charged particle approaches the Klein-Nishina regime ($\gamma E_{\gamma,\mathrm{le}} \approx mc^2$ in the lab frame), the energy of the scattered photon will be increasingly reduced with respect to that predicted by Eq. 2.29, until for extremely high particle energies $E = \gamma mc^2$, the mean scattered photon energy can be expressed in the from

$$< E_{\gamma,\mathrm{he}} > = \frac{1}{2} < E > \tag{2.30}$$

This is also called the *Klein-Nishina limit*.

2.3 The blazar spectrum

When plotted as frequency times differential energy flux (νF_ν) *versus* frequency ν – or equivalently, energy squared times particle flux ($E^2 \mathrm{d}N/\mathrm{d}E$) *versus* energy E – the spectral energy distributions (SEDs) of all observed blazars, as well as other astrophysical objects producing energetic non-thermal emission such as supernova remnants, consist of two distinct broad peaks: a broad hump at lower energies followed by a similarly

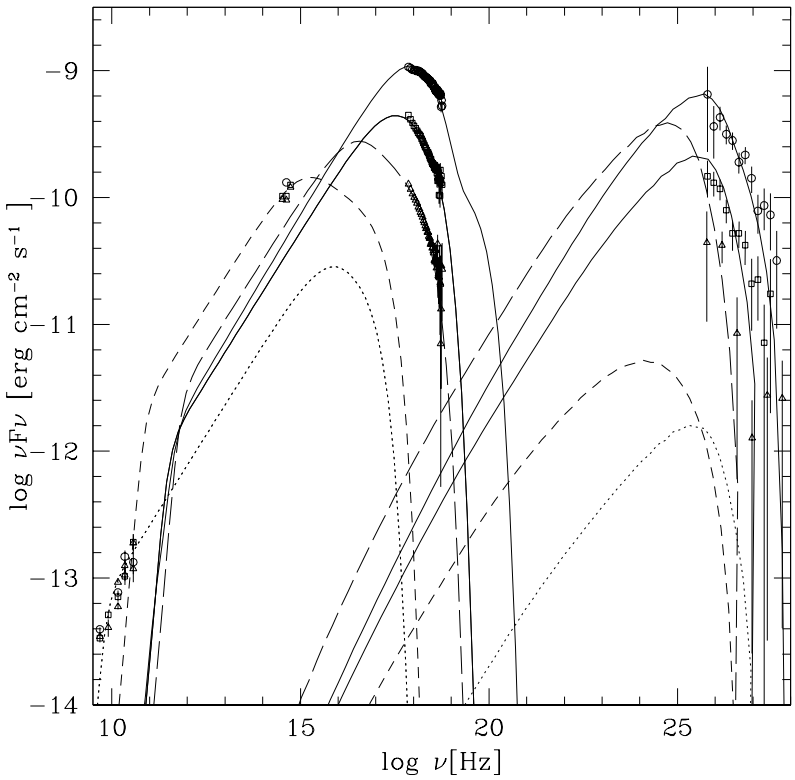

Figure 2.7: The SED of the nearby HBL-type blazar Mrk 421, showing X-ray fluxes from the RXTE-PCA spectrometer and simultaneous VHE gamma-ray data from the Whipple ACT, for three different flux states (from low to high: *triangles, squares* and *circles*). A separate SSC model fit is shown for each of the flux states (*solid* and *long-dashed* lines). To fit the highest flux state, a two-zone SSC model was required. The dashed and dotted lines represent contributions from additional zones that were introduced to also fit the simultaneous radio and optical data shown. Notable is the higher IC-peak flux predicted by the low-state fit compared to that of the medium-state fit. Figure taken from Blazejowski *et al.* (2005).

shaped, partially overlapping hump at higher energies. The SED of the well studied blazar Mrk 421 is shown in Fig. 2.7, including X-ray and gamma-ray measurements as well as some emission model fits which will be explained later. The double-humped structure of the energy spectrum can be seen clearly.

For reasons that will be addressed in Sec. 2.4.3, the absolute *energy scale* of the spectra can vary dramatically from one blazar to another. The blazars that interest us here are those for which the high-energy bump reaches up to VHE gamma-ray energies.

Observationally, one particular detector can access only a small band of the total spectrum of non-thermal emission from blazars, which covers up to 20 orders of magnitude in energy from radio up to X-ray or VHE gamma-ray energies. Over sufficiently

small intervals of the spectrum and especially on the falling edges of the two peaks (*i.e.* on the high-energy slopes), the spectra can be well fitted by simple power laws, which means that within these energy ranges, the differential energy flux (F_ν) or differential particle flux ($\mathrm{d}N/\mathrm{d}E$) can be described by means of the simple expression

$$F_\nu \equiv \frac{\mathrm{d}F}{\mathrm{d}\nu} = F_0\,\nu^{-(\alpha-1)} \tag{2.31}$$

$$\frac{\mathrm{d}N}{\mathrm{d}E} = N_0\,E^{-\alpha} \tag{2.32}$$

The quantity α is called *spectral index* and is generally in the range of $[0,\,5]$.[13]

In the following sections we will examine the two distinct components of the blazar SED in more detail, and give an introduction to the two most popular groups of models for gamma-ray emission in the relativistic jets of blazars. For more details on the emission spectra of blazars, the reader is referred to Padovani (1995) and Costamante, Ghisellini *et al.* (2001).

Synchrotron peak

In all current models, the low-energy hump in the blazar SED is caused by synchrotron radiation from electrons (and possibly positrons), which are accelerated – presumably by means of first and second order Fermi acceleration (see Sec. 2.2.3) – within the magnetic fields of the jet. This produces a power law distribution of ultra-relativistic accelerated electrons/positrons. At the low-energy end, any turn-on feature of the injected spectrum of accelerated electrons/positrons will be mirrored in the synchrotron spectrum; in addition, a further break is introduced by synchrotron self-absorption (see Sec 2.2.4), which sets in when the energy drops below a certain threshold that is determined by the energy density in the emission region. As the energy increases, the (presumed) power law spectrum of accelerated particles with spectral index p causes the emission of synchrotron radiation in the form of a power law with spectral index as prescribed by Eq. 2.18. At some *break energy* E_{br}, radiative cooling – mainly due to synchrotron emission – becomes more efficient than the adiabatic energy loss due to particles escaping the shocks (and generally the acceleration regions of the jet). The result is a break in the spectrum of the accelerated particles at E_{br}, beyond which the spectrum softens. This spectral break is also mirrored in the synchrotron spectrum emitted by the electrons/positrons. Geographically, the bulk of the synchrotron emission occurs in the upstream part of the jet, where the magnetic fields are highest.

[13] Care needs to be taken in regard to the exact definition of α. In the literature, the spectral index as it is defined here is sometimes called Γ instead of α, in which case α is often defined as $\alpha \equiv \Gamma - 1$.

High-energy peak

Owing to less efficient observation techniques at the higher energies – particularly at gamma-ray energies – and because of the exponential flux decrease with increasing energy, data are much more sparsely sampled in the range of the high-energy peaks (from now on *HE-peaks*) of the SEDs of most observed blazars. Consequently, theoretical emission models cannot be constrained yet in a satisfactory way. As of now, several different mechanisms are invoked to explain how exactly the HE-peak is produced. The different approaches shall be described in the following two sections. Independent of the exact production mechanism, however, the observed spectra of blazars whose HE-peaks extend into the VHE gamma-ray band may be cut off at the high-energy end due to gamma-ray absorption in the extragalactic radiation fields. This effect will be addressed in Sec. 2.5.

2.3.1 Leptonic jet models

Building on the convincing evidence for the acceleration of charged particles within relativistic jets, the final question to be addressed by the jet models is how gamma rays, or in general the emitted photons of the high-energy peak, are produced in the jets, and by which types of particle. There are currently two major groups of models, differing in the composition of the accelerated matter in the jets.

In the so-called *leptonic* emission models, gamma rays are produced through Inverse-Compton (IC) scattering of low-energy photons by accelerated electrons, and possibly positrons.[14] As motivated in the previous sections, there exists convincing evidence that the low-energy peak in the blazar SED is caused by synchrotron radiation from electrons accelerated within the magnetic field of the jet. From the energy scale of the synchrotron spectra, it is known that the emitting electrons must possess very high energies, high enough to Compton-upscatter any ambient low-energy (radio to X-ray energy, the exact range depending on the source) photons to hard X-ray and gamma-ray energies via the IC process (see Sec. 2.2.5 for details). From IC-kinematics and what was said about synchrotron emission in the previous section, it follows that these upscattered photons roughly "mirror" the shape of the broad synchrotron peak, providing an intuitive explanation of the HE-peak. The latter is therefore often also called *Inverse-Compton peak* (IC-peak), which is already hinting at the fact that the leptonic scenario is currently the preferred gamma-ray production model for blazars.

[14] In the following, the term "electron" should be understood to refer to both electrons and positrons, even though the presence of positrons within astrophysical jets has not yet been confirmed.

Synchrotron Self-Compton

Depending on the presumed source of the low-energy "seed" photons, leptonic models are further classified into two sub-categories. In the generally favoured *Synchrotron Self-Compton* (SSC) model, the seed photons are the very synchrotron photons which are emitted by the relativistic electrons themselves (hence the term "self"). This idea, which can be traced back to Jones, O'dell & Stein (1974), was first used by Maraschi, Ghisellini & Celotti (1992) as a basis for a model that successfully explained the double-humped broad band SED of the blazar 3C 279. In the meantime more refined SSC models have been developed, and things such as time-dependent emission have been included in order to explain the observation of flaring behaviour (Mastichiadis & Kirk, 1997).

The simplest, yet often adequate form of the SSC model follows the prescription of Tavecchio, Maraschi & Pian (2001) and has been used to successfully model the spectra of many blazars. This model assumes a single emission region threaded with a tangled magnetic field of uniform strength, and depends only on a limited number of parameters. The free parameters are listed in the following, including typical fit values for observed VHE gamma-ray emitting blazars (*i.e.* BL Lacs):

- size of the emitting volume R, which is usually assumed to be spherical or cylindrical ($R \approx 10^{14}$ - 10^{17} cm)
- number density (K) or energy density (ρ_e) of the relativistic electron population, which together with R is responsible for the normalisation of the emission spectrum ($K \approx 10^{3}$ - 10^{10} cm^{-3})
- magnetic field B of uniform strength and random orientation, which determines the luminosity of the synchrotron component ($B \approx 0.01$ - 1.5 G)
- bulk Lorentz factor Γ of the emitting volume ($\Gamma \approx 6$ - 30)
- jet viewing angle θ; Γ and θ determine the Doppler factor δ and thus affect the energy scale of the emitted spectrum ($\delta \approx 10$ - 40)
- two energy intervals describing the injected electron population, given by the lower and upper bounds $\gamma_{\rm min}$, $\gamma_{\rm max}$ on the Lorentz factors of the injected particles,[15] and a *break energy* $E_{\rm break} = \gamma_{\rm break} m_e c^2$ at which the spectrum softens ($\gamma \in [0, 10^8]$)
- spectral index p of the accelerated electron population ($p \approx 1.6$ - 3.0)

[15] Note that these Lorentz factors $\gamma = E/m_e c^2$ are referring to the rest frame of the acceleration region within the relativistic jet, which itself is moving with the bulk Lorentz factor Γ.

Here $E_{\rm break}$ is the energy at which radiative cooling begins to exceed the intrinsic escape rate of the accelerated electrons, manifesting itself as an increase of the electron spectral index by one power, *i.e.* from p below $E_{\rm break}$ to $(p+1)$ above it.[16] In most SSC models the spectral softening is constrained to exactly 1.0, but some models allow it to vary within a certain range. The large spread in the size of the emission region is unlikely to be physical, in fact most authors seem to obtain fits with $R \approx 10^{15}$ cm. Fig. 2.7 shows examples of SSC model fits to the SED of a well studied blazar (Mrk 421) for different flux states (see also Sec. 2.3.3). It should be noted that radio fluxes are not normally used to fit SSC models, since they include a thermal component.

In the SSC model, B cannot be too high; in most blazars it is fitted to be below 1 G.[17] Since the synchrotron intensity increases with the square of the magnetic field strength (see Eq. 2.17), too high a magnetic field density will result in so many synchrotron photons that the energy loss of the relativistic electrons becomes too large to sustain a highly relativistic electron population. In this case the electrons lose energy by synchrotron radiation and IC-scattering on synchrotron photons, but also on already IC-scattered photons ("Compton catastrophe"). One fundamental quantity in the SSC model is the synchrotron peak frequency $\nu_{\rm peak}$, as it is often directly measurable, and it also determines the peak energy of the IC emission. From Eq. 2.16 we know that $\nu_{\rm peak}$ is proportional to $B\gamma_e^2$, and from Eq. 2.18 with $p = 2\text{-}3$ it follows that γ_e must in this case be the maximum Lorentz factor of the injected electron spectrum, $\gamma_{\rm max}$.[18]

The energy corresponding to the IC-peak can be predicted by folding the broken power law that describes the accelerated electron energies with the summed synchrotron spectrum (*i.e.* the "hump"). To first order – although strictly only valid for monoenergetic electrons – the energy at the IC-peak can be approximately forecast from the synchrotron peak energy by applying the squared Lorentz factor $\gamma_{\rm break}^2$ of the electrons at the break energy according to Eq. 2.29. For blazars whose IC-peaks reach up to GeV energies, however, the IC-flux becomes suppressed by Klein-Nishina effects (see Sec. 2.2.5), causing a steeper-than-normal decline of the flux from $\sim$1 TeV on and thus invalidating

[16] These spectral indices are obtained from approximate solutions of the differential equation for the electron density's time-evolution.

[17] Details on magnetic field strengths in the SSC model for various blazar types are given in Sec. 2.4.3.

[18] This becomes evident when one recalls that the SED is normally plotted (double-logarithmically) as flux times energy squared over energy, *i.e.* νF_ν versus ν (or equivalently, $E^2\,{\rm d}N/{\rm d}E$ *vs.* E). Hence the synchrotron spectrum, which has a spectral index of less than 2 even after the break, will only start to fall off again for energies greater than those resulting from the highest-energy injected electrons, *i.e.* those with energy $\gamma_{\rm max} m_e c^2$.

this prescription.[19] The relativistic motion of the emission region itself causes all emitted radiation to be further Doppler-boosted, and if the redshift of the source is significant, observed frequencies will be redshifted as well. We can thus estimate the *observed* synchrotron peak frequency from integrating Eq. 2.16 over pitch angle and applying the Doppler factor and cosmological redshift correction, using the above arguments for the value of γ:

$$\nu_{\rm peak} = \frac{3}{2\pi^2} \frac{\delta}{1+z} \frac{e\,B\,\gamma_{\rm max}^2}{m_e\,c} \tag{2.33}$$

External Compton

Apart from synchrotron photons produced by the jet itself, the soft target photons may also be ambient radiation from other sources. Dermer, Schlickeiser & Mastichiadis (1992) proposed that the IC-target photons come from sources *external* to the jet. This concept is known as the *External Compton* (EC) model.[20] Sources of the low-energy photons may be direct thermal radiation from the accretion disk or hot electron corona, as well as scattered radiation from the broad-line region or even the dusty torus. The photons scattered in the broad-line clouds may come from the disk, the corona or the jet itself.

In general – assuming that gamma rays are indeed produced by inverse Compton scattering of electrons – there is no physical reason why both of these mechanisms should not operate in blazar jets at the same time. Principally, one should be able to distinguish between contributions from the SSC and EC mechanisms within a particular source by measuring the time delay between flares in the synchrotron component to the emergence of the flares in the inverse-Compton peak. In the SSC model no significant time delay should be measurable, as the synchrotron and IC emission would occur in the same volume, whereas in the case of an external photon source, the IC-photons would be produced at a slightly later time.

In the "pure" SSC model, the flux at the IC-peak is constrained by the synchrotron flux; for magnetic field values commonly required in leptonic models, its value of νF_ν can only be as high as that at the synchrotron peak, or lower. Some blazars however have broad-band SEDs with higher IC fluxes than that, and are thus not well fitted by an SSC model alone.[21] In such cases assuming the presence of a population of external

[19] As we will see in Sec. 2.5, the spectrum gets further distorted by gamma-ray absorption in the extragalactic medium on the way from the source to the observer, which also causes a gradual cutoff of the spectrum above $\sim$ 100 GeV, effectively concealing the Klein-Nishina suppression for all but the closest blazars.

[20] Sometimes this is also referred to as the *External Inverse Compton* (EIC) or *Synchrotron External Compton* (SEC) model.

[21] This is usually the case for the SEDs of FSRQ, which will be introduced in Sec. 2.4.1.

target photons in the emission region often provides a satisfactory fit of the data, as increasing the external photon density will increase the IC-flux without affecting the synchrotron component. Informative plots showing the contributions of various target photon sources in a combined SSC/EC model fit to different blazar SEDs can be found *e.g.* in Hartman *et al.* (2001). In the following, we will refer to models that include contributions from both SSC and EC mechanisms as "SSC/EC" models.[22]

The simplest types of leptonic models, such as the one described above, are so-called *single-zone* models. In these, the emission region is assumed to be one homogeneous blob of plasma, moving with a constant Lorentz factor Γ. In some cases, fits can be improved by simply including additional zones with separate fit parameters (see *e.g.* Blazejowski *et al.*, 2005); these are called *multi-zone* models. Another step up in level of detail is achieved by assuming inhomogeneous emission regions, examples of which can be found in Blandford & Levinson (1995); Levinson & Blandford (1995); Levinson (1996). Instead of postulating a uniform jet velocity, more realistic models may also be obtained by allowing for a velocity structure within the jet (Ghisellini, Tavecchio & Chiaberge, 2005).

2.3.2 Hadronic jet models

As was motivated in Sec. 2.2.3, in some jet models charged hadronic particles are accelerated along with electrons to ultra-relativistic energies. As an alternative to the leptonic gamma-ray emission models, a number of promising hadronic models have therefore been developed, which also have been able to explain the SEDs of blazars. In these types of models, protons (and to a lesser degree, heavier ions) are accelerated along with electrons (and possibly positrons) within the relativistic jets, by means of the same acceleration mechanisms. As in the leptonic models, the electrons/positrons produce synchrotron radiation in the magnetic field, and are thus responsible for much of the low-energy peak of the blazar SED. The HE-peak however is caused by secondary particles resulting from the acceleration of hadrons, while gamma rays produced by *direct* IC-scattering of accelerated electrons/positrons pair-produce on the high density of low-energy photons and thus initiate electromagnetic cascades. The latter is a result of the much stronger magnetic fields required by hadronic models, which do not allow for a highly relativistic electron population to develop *inside* the emission volume (see previous section). Note

[22] In the literature, these models are sometimes confusingly called "EC", even though they do include SSC as well.

that the reason why hadrons do not IC-scatter with photons directly becomes evident when recalling that the cross-section for IC is proportional to m^{-2} (see Eq. 2.27).

The first gamma-ray emission model based on the acceleration of protons was developed by Mannheim (1993). This model relies exclusively on the hadronic cascades (also called *proton-induced cascades* or PIC) that are produced when the accelerated ultra-relativistic protons scatter inelastically with ambient low-energy photons. These target photons can, just like in the leptonic models, be either synchrotron photons emitted by the relativistic electron population or, to a lesser degree, photons from external radiation fields. At the extremely high threshold energy of $\sim 10^{20}\,\mathrm{eV}$, protons will undergo photoproduction of highly relativistic mesons and $\mathrm{e}^+/\mathrm{e}^-$ pairs:

$$\begin{aligned} p + \gamma &\longrightarrow p + \pi^0 \ (+\ \pi^+\ \pi^-\ \pi^0) \\ &\longrightarrow n + \pi^+ \ (+\ \pi^+\ \pi^-\ \pi^0) \\ &\longrightarrow p + e^+ + e^- \ (+\ \ldots) \end{aligned}$$

Apart from pions, kaons and heavier mesons may also be produced. The electron-positron pairs generate synchrotron radiation in the magnetic field just like the directly accelerated electrons, adding to the low-energy peak of the blazar spectrum. They then up-scatter soft photons via Inverse Compton, as in the leptonic models, producing secondary gamma rays. The difference is that in hadronic models, the ambient photon densities are so high that the emission volume is opaque to gamma rays of the highest energies, which pair-produce and initiate electromagnetic cascades instead. Gamma rays are only able to leave the emission region when the energy of the secondary gamma rays has dropped below a threshold energy for which the optical depth decreases below unity. Gamma rays produced close to the edge of the emission volume are also able to escape, resulting in the highest energy radiation originating from a thin surface layer of an optically thick pair plasma, while lower energy gamma rays and synchrotron photons may come from anywhere within the emission region.

The secondary mesons do not escape the jet, but decay almost instantly (π^0) or after a short time ($\pi^\pm$, $K^\pm$, ...)

$$\begin{aligned} \pi^0 &\longrightarrow \gamma + \gamma \\ \pi^+ &\longrightarrow \mu^+ + \nu_\mu \\ \pi^- &\longrightarrow \mu^- + \bar{\nu}_\mu \end{aligned}$$

Gamma rays from neutral pion decay initiate electromagnetic cascades. Muons decay after a mean time of $2.2 \cdot 10^{-6}$ s in their rest frame to positrons and electrons, which cascade further downstream of the initial vortex:

$$\begin{aligned} \mu^+ &\longrightarrow e^+ + \nu_e + \bar{\nu}_\mu \\ \mu^- &\longrightarrow e^- + \bar{\nu}_e + \nu_\mu \end{aligned}$$

Note that ample neutrinos are produced in most of these processes. An unambiguous indicator for the acceleration of hadrons within relativistic jets would thus be the detection of UHE neutrinos from the direction of blazars.

More recent hadronic models such as the *Synchrotron Proton Blazar* (SPB) model of Mücke & Protheroe (2001) also include VHE gamma-ray contributions from synchrotron radiation produced by the accelerated protons, as well as from charged secondary particles (muons and, due to their shorter life times in a less significant way, charged mesons such as $\pi^\pm$ and $K^\pm$). While the cascades give rise to a rather continuous gamma-ray spectrum, the synchrotron radiation produces spectra similar to that of the IC-peak from electrons, although slightly more peaked. In blazars with low radiation densities (HBLs), the proton synchrotron emission dominates over the cascade contribution, as photoproduction in these objects is highly inefficient. Likewise, synchrotron radiation from secondary particles is expected to produce an important component to the SEDs of FSRQ and LBLs, but is insignificant in HBLs.[23] The SPB model provides a good description of observed blazar spectra (Mücke, Protheroe *et al.*, 2002).

Fitted parameters such as size of emission region, bulk Lorentz factor and jet viewing angle are similar in hadronic and leptonic models. One key difference between the two types of models is the magnetic field strength. While the latter can not be significantly larger than ~ 1 G in leptonic models (see previous section), the high energy threshold for photoproduction requires protons to be accelerated to extremely high energies. This can only happen as long as the proton Larmor radius does not exceed the radius of curvature of the shocks within the jet (Mannheim, Westerhoff *et al.*, 1996). In order for protons to reach the required energies, a minimum magnetic field strength of 10 G is necessary within the emission region (Mücke, Protheroe *et al.*, 2002), and generally values between $B \sim 30$ - 40 G or even higher are assumed. Another important prerequisite for hadron acceleration to occur is that the proton acceleration time scale is shorter than the expansion time scale of the jet. Given that these conditions are fulfilled and hadrons are accelerated efficiently, the "advantage" of hadronic models is that, in theory,

[23] These different types of blazars will be described in Sec. 2.4.3.

protons can be accelerated to much higher energies than electrons/positrons, as they suffer significantly less losses through synchrotron radiation. Furthermore, electrons suffer absorption by IC-scattering in regions of high photon densities, whereas protons may escape those regions and photoproduce in areas of decreased radiation density. As such, hadronic models are able to sustain gamma-ray emission at much higher energies than leptonic models. An upper limit on the energy to which protons can be accelerated may be the threshold for photo-pion production, *i.e.* $\approx 10^{20}$ eV. Finally, protons should be able to be accelerated more efficiently within shocks than electrons, as the Larmor radius of electrons is much smaller and the shocks have a finite thickness, preventing some low-energy electrons from traversing one of the two shock fronts. This seems to be supported by the cosmic ray composition, in which only about 2% of the energy is due to electrons, and the rest hadrons (Schlickeiser, 2003). This argument is valid under the assumption that cosmic rays are indeed produced in astrophysical jets.

Mannheim, Westerhoff *et al.* (1996) successfully fitted the spectra of nearby blazars with the pure cascade model, and agreement with the data appears to be further improved when the p and μ synchrotron radiation is included (*e.g.* Böttcher, Mukherjee & Reimer, 2002). Obviously, due to the much larger number of physical processes contributing to the emission, hadronic models are much more complex than their leptonic counterparts. This also means that finding the right set of parameters to fit a particular blazar spectrum is inherently more difficult. Due to the simplicity of leptonic models and the fact that they seem to describe most blazar spectra very well, they are often preferred, particularly among experimentalists. The high magnetic fields required by hadronic models, and the still uncertain answer to the question whether astrophysical jets really do accelerate hadrons to the required high energies, may be other reasons. Nevertheless both types of models seem to successfully fit most blazar SEDs (see *e.g.* Böttcher & Reimer, 2004).

The case for hadronic jet models is strengthened by the fact that these models provide for a natural solution to the unsolved mystery of the origin of ultra-high-energy cosmic rays (see Sec. 1.3). Photoproduction of charged pions by protons gives rise to ultra-relativistic neutrons, which carry about 4/5 of the original proton energy. These neutrons escape the magnetically confined acceleration region freely, and due to time dilation the highest energy neutrons also escape the host galaxy and its (weaker) magnetic fields before β-decay occurs. The luminosity of the protons from neutron decay is hence comparable to the gamma-ray luminosity, and might be able to account for the extragalactic cosmic-ray proton flux above $10^{18.5}$ eV (Hayashida *et al.*, 1996; Mannheim, Westerhoff *et al.*, 1996).

Since hadronic processes also give rise to very high neutrino fluxes, a detection of UHE neutrinos from the direction of a blazar by one of the cubic-kilometre scale neutrino detectors currently under development (Sec. 1.4) would provide strong evidence for the presence of hadron acceleration within relativistic jets, and thus of hadron-initiated gamma-ray emission and UHE cosmic ray acceleration in blazars.

2.3.3 Flux variability

As mentioned earlier, emission from blazars has been observed to be highly variable across all energies. In order to be able to constrain emission models it is therefore important to use *contemporaneous* flux measurements at various different energies. Therefore, as well as for studying correlations between different parts of the SED, so-called *multiwavelength campaigns* are necessary, during which a particular source is monitored simultaneously in two or more frequency bands for longer periods of time (usually of the order of weeks). Through such campaigns it was found that in many instances, synchrotron and high-energy fluxes of blazars are highly correlated (see *e.g.* Blazejowski *et al.*, 2005).[24] More precisely, the energies at which the synchrotron and HE distributions peak as well as the absolute flux values of the two distributions are correlated, whereby higher peak energies correspond to higher fluxes and vice versa. Depending on whether the flux values are low or high (in relative terms), the blazar is said to be in a *low flux* or a *high flux state*. VHE gamma-ray emission from blazars that normally do not emit detectable fluxes in that energy band has been discovered during optical outbursts of the source in more than one instance (Albert *et al.*, 2007b). The SSC model in particular provides a convenient explanation of the observed correlations, since synchrotron emission and IC-emission are due to one and the same population of relativistic electrons. This is demonstrated in the successful SSC fits to the different flux states of Mrk 421 shown in Fig. 2.7. Whether these correlations could be used to rule out a hadronic origin of the observed gamma rays is questionable, as hadronic models would also give rise to correlations if the photoproduction targets are synchrotron photons. In all discussed emission models, variability may be due to a number of reasons. Short-term changes within the shocks or magnetic reconnection processes may play a role; on longer time scales, changes in the propagation direction of the emitting regions, maybe due to jet precession, might alter the degree of Doppler boosting.

[24] For most blazars that have been detected in the gamma-ray band, this translates into a correlation between optical/X-ray and gamma-ray fluxes.

One interesting type of flux variability in blazars in as far as gamma-ray emission models are concerned is the occasional observation of so-called *orphan flares*, which are flares that are seen in the HE-peak (usually in gamma rays) but not in the synchrotron emission (Blazejowski *et al.*, 2005). It has sometimes been suggested that orphan flares may indicate the presence of hadronic processes within the jet, as in hadronic models, synchrotron and IC emission do not necessarily originate in the same physical volume. However, this argument is only valid in the case where the simple one-zone SSC model, which indeed appears to be unable to account for such flares, is the only alternative scenario. In reality there are multiple ways to accommodate orphan flares in leptonic models, such as by assuming multi-component SSC models, certain orientations of the local magnetic field with respect to the line of sight, or a significant contribution from EC (in which case target photon densities may vary temporally as well as spatially independently of the synchrotron emission). Furthermore, there are examples of orphan flares where hadronic models are considered unlikely to be the cause (Krawczynski *et al.*, 2004).

2.4 A unified model of AGN

Up to now, our treatment of AGN has been focussed on a general model of these objects (Sec 2.1) as well as on a variety of physical processes within AGN, in particular within relativistic jets (Sections 2.2 and 2.3). It has also been explained that observationally, AGN manifest themselves in a variety of forms, part of which is due to different viewing angles with respect to their axis of rotation. In this section we will explore this concept further, and attempt to develop a unification of all observed types of AGN. A classification of AGN in general will be presented first, after which the sub-class of blazars will be examined in more detail.

2.4.1 AGN classification

The current paradigm for AGN presented in Sec. 2.1.3 consists of an accreting SBH whose accretion disk emits a bright continuum of thermal and non-thermal radiation. The clouds of gas rotating around the SBH on either side of the accretion disk (broad-line and narrow-line emission regions) as well as the obscuring torus are also relevant for the following classification. The existence of bipolar relativistically outflowing jets that propagate along the SBH's axis of rotation is confirmed in the case of radio-loud AGN, where they can often be resolved via prominent radio emission from the jets or from large radio lobes. Only around 20% of AGN are radio-loud; the remaining AGN,

which display little to no radio emission, may also possess jets, but ones which are weak enough to evade detection. There seem to be some inherent differences between radio-loud and radio-quiet AGN which cannot be explained by selection effects, hence these two sub-categories need to be treated separately. We will first turn our attention to the radio-loud AGN, which are more important in the context of this work.

An initial classification of the various types of radio-loud AGN was undertaken by Urry & Padovani (1995). Building on this work, most observational differences between AGN can be explained by a variation of the *jet viewing angle*, *i.e.* the angle between jet axis and line of sight. Given a particular AGN, different properties of the object can be observed depending on the angle at which the relativistically beamed jet is seen. The following categories of radio-loud AGN are distinguished by progressively decreasing jet viewing angles:

Narrow-line radio galaxies

Radio-loud AGN with jets oriented at relatively large angles to the line of sight are observed on Earth as radio galaxies. The jets are easily identifiable due to their strong radio emission. Since emission is not enhanced by Doppler effects, all resolved radio galaxies are relatively close. For large viewing angles, the broad-line region close to the accretion disk is usually obscured by the dusty torus, hence no broad-line emission is observed. The spectra of these radio galaxies only contain narrow lines emitted by the slower moving clouds further away from the central black hole, hence they are called *narrow-line radio galaxies* (NLRG). Fanaroff & Riley (1974) classified radio galaxies according to their radio morphology into those with core dominated (*Fanaroff-Riley* type I, short FR I) and those with jet dominated (FR II) radio patterns. FR II galaxies show very straight jets out to hundreds of kpc from the black hole; they appear to have high Lorentz factors and usually terminate in hot spots within large radio lobes. Their integrated radio luminosity is much higher than that of FR I galaxies, with the majority of radio emission originating in the radio lobes. The jets of FR I radio galaxies are less powerful and become warped some distance away from the nucleus. Their intensity also falls off with increasing distance from the SBH. They are of lower total intensity than FR II's, but radio intensities emitted by the (upstream parts of the) jets are comparable. Both types of radio galaxies are thought to form a single continuous distribution, as there appear to be radio galaxies that show some characteristics of either subclass.

Broad-line radio galaxies

As the jet viewing angle decreases, emission from the broad-line clouds gradually becomes visible, *i.e.* the optical spectra of these *broad-line radio galaxies* (BLRG) contain narrow as well as broad emission lines. Radio galaxies that possess no or only a thin torus outside of the accretion disk may appear as BLRG even for larger viewing angles.

Quasars

The AGN known as *quasars* are intrinsically more luminous than BLRG, and at the same time more distant; the jet viewing angles probably overlap with those of BLRG. Relativistic beaming effects due to the approaching jet increase with decreasing jet viewing angle; when jets are visible, superluminal motion can sometimes be seen. Doppler boosting combined with the large redshifts of all observed quasars results in their host galaxy being de facto invisible (hence the term "quasi-stellar"). Like those of BLRG, quasar spectra contain broad and narrow emission lines. Having intermediate to small viewing angles, quasars are further divided into two sub-categories: *Flat Spectrum Radio Quasars* (FSRQ) exhibit relatively flat (hard) radio spectra with index $\alpha < 0.5$, while *Steep Spectrum Radio Quasars* (SSRQ) show softer spectra ($\alpha > 0.5$). Both types of quasars are thought to form a continuous sequence, with the distinction arising from smaller viewing angles (and therefore larger Doppler factors) in the case of FSRQ when compared to SSRQ. A prominent feature of all quasar spectra is the conspicuous UV-excess ("blue bump"), which is most likely thermal radiation from the hot accretion disk.

Blazars

In blazars, the jets are oriented within a few degrees ($\lesssim 6°$) from the line of sight, giving rise to strong relativistic beaming effects, consisting of rapid variability and Doppler-enhanced energy and luminosity. Due to the extreme jet dominance of the emitted radiation, it has a high degree of polarisation (from synchrotron emission), and no blue bump from the accretion disk is visible. Because they are seen almost head-on, the jets of blazars are often not identifiable as such, and the strong non-thermal emission is highly core dominated.

Blazars can be subdivided into so-called *Optically Violently Variable* Flat-Spectrum Radio Quasars (OVV) and *BL Lacertae* (BL Lac) objects. OVV, most commonly just referred to as FSRQ, are the continuation of the above mentioned FSRQ quasars for progressively smaller viewing angles, giving rise to increasing spectral hardness and

variability (hence the term "OVV"). In contrast to FSRQ, BL Lac objects are characterised by nearly featureless optical spectra, *i.e.* an apparent absence of or only very weak emission lines, and a high degree of linear polarisation (up to $P \sim 30\%$) in infrared to X-ray wavelengths, which is also highly variable (Januzzi, Smith & Elston, 1992). This class of blazars is named after the first object that was found to have the described properties, BL Lacertae (1ES 2200+420).

Although the SEDs of FSRQ and BL Lacs are similar in shape, they differ in terms of energy scale and relative luminosity. BL Lacs are overall less luminous, and their synchrotron and HE-emission components peak at higher energies than those of FSRQ. Furthermore, in FSRQ the bolometric luminosity of the high-energy peak dominates over that of the synchrotron peak, whereas in BL Lacs this is not necessarily so. In Sec. 2.4.3 it will be shown how the different types of blazars fit into a unified AGN model.

Since they can have a very wide range of synchrotron peak energies, BL Lacs are sub-classified into so-called *Low-energy* and *High-energy peaked BL Lacs*, LBL and HBL, according to the spectral location of the synchrotron peak; for LBLs the synchrotron peak is in the infrared or optical bands ($\nu_{peak} < 10^{15}$ Hz), whereas BL Lacs with synchrotron peaks in the UV or soft X-rays ($\nu_{peak} \geq 10^{15}$ Hz) are referred to as HBLs.[25]

AGN unification

There is now a general consensus that FR I radio galaxies are the parent population of BL Lacs, *i.e.* BL Lacs are FR I galaxies whose jet axes are aligned closely with the line of sight (see *e.g.* Bai & Lee, 2001). These objects are thought to form a single population of low-luminosity, radio-loud AGN. Likewise, there is increasing evidence that FR II galaxies are associated with FSRQ, forming a subclass of AGN with powerful jets and high luminosities. In both cases, the rather sharp divisions between each of these subclasses (low *vs.* high jet luminosity, no beaming *vs.* strong beaming) seen in early observations are now believed to be the result of selection effects; as more and more objects are examined in detail, evidence is increasing that the properties of AGN do indeed change continuously from one extreme to the other. However it also appears that the parameters used to distinguish the various AGN classes are not very well defined. In a recent study by the VLA radio telescope, Heywood *et al.* (2007) found that there are examples of FR I galaxies which can also be associated with quasars, and that some radio

[25] Initially BL Lacs were classified observationally according to the ratio of X-ray to radio flux, distinguishing between *radio-selected* (RBL) and *X-ray selected* (XBL) BL Lacs. In the meantime this scheme has been replaced by the division into LBL and HBL, which is more physical.

galaxies with FR I morphology have a radio luminosity *above* the traditional FR I/FR II break luminosity.

Radio-quiet AGN

As mentioned above, radio-quiet AGN are believed to have either very weak jets, or none at all. Like their radio-loud counterparts, they can be classified according to viewing angle – in this case not referring to the jet axis, but the axis of rotation of the SBH and accretion disk – and luminosity. The least luminous radio-quiet AGN are the well known *Seyfert* galaxies. At first sight, Seyferts do not appear very different from "normal" (inactive) galaxies, except they have an optically bright core and give rise to optical line emission like radio galaxies. As in the case of radio galaxies, there are two types: Seyfert II galaxies are thought to be viewed more or less edge-on and are thus the radio-quiet counterparts of NLRG, *i.e.* obscuration by a cold torus allows only narrow emission lines to reach the observer. Seyfert I galaxies, presumably counterparts of BLRG and viewed at larger angles to the plane of the disk, are observed to emit both narrow and broad lines. Type-I Seyfert cores are more luminous at optical wavelengths than those of Type-II Seyferts, likely also due to obscuration effects. One curious fact about Seyferts is that they are predominantly – but not exclusively – spiral galaxies, while all other types of AGN appear to be exclusively harboured by large elliptical galaxies.

The high-luminosity version of Seyfert II galaxies is not yet well established, presumably because they are too distant and too faint. Some objects that are bright in the infrared and appear to show Type II optical emission, *IR-* or *IRAS-quasars*, have been proposed to fill in the "gap" (for references, see *e.g.* Urry & Padovani, 1995). More luminous radio-quiet AGN with smaller viewing angles, and hence bright enough to be identified as Type I, are called *Quasi-Steller Objects* (QSO). These are the counterparts of quasars, and like the latter they are too far for the host galaxy to be resolved.[26] There is no certainty as to which objects are represented by QSO whose jets are viewed head-on; one candidate are so-called *BAL quasars* (Broad Absorption Line quasars). The radio-quiet counterparts of BL Lacs, if they exist, are unknown. An overview of all AGN types discussed in the preceding is given in Table 2.1.

[26] Actually, the original definition of "QSO" includes *both* radio-quiet and radio-loud objects (Carrol & Ostlie, 1996), but in the literature this term is often used to refer only to the radio-quiet type, which is what will be done here as well.

	Type 2 (Narrow lines)	Type 1 (Broad lines)	Type 0 (Unusual)	Lumin.
Radio-quiet	IR Quasar Seyfert II	QSO Seyfert I	BAL-QSO ? ?	*high* *low*
Radio-loud	FR II FR I	SSRQ FSRQ BLRG	FSRQ (OVV) BL Lac	*high* *low*
$\Longrightarrow$ *decreasing angle between jet axis and line of sight* $\Longrightarrow$				

Table 2.1: Classification of Active Galactic Nuclei based upon observational properties, ordered horizontally by jet viewing angle and vertically by radio-loudness (left column) and relative luminosity within each category (right column). FR I and FR II radio galaxies are *narrow line radio galaxies*; SSRQ and FSRQ are *quasars*, *i.e.* radio-loud QSO; OVV-FSRQ and BL Lacs are *blazars*. Adapted from Urry & Padovani (1995).

2.4.2 AGN activity

Having simplified the zoo of observationally different AGN types to a smaller set of classes which are distinguished almost exclusively by the power of their jets, the ultimate question regarding a self-consistent AGN model remaining to be answered is: what exactly causes some AGN to have strong jets, and others none? This may relate to the more fundamental question of what causes some galaxies to be active and others not, provided all (or most) galaxies harbour a SBH at their centre.

The reason for the activity in some galaxies has been one of the key questions in AGN research. It has been suggested that close encounters with companion galaxies may trigger activity, and indeed some BL Lacs seem to have companion galaxies showing signs of tidal interactions (Stickel *et al.*, 1993). Whether this can explain the activity in all types of AGN is doubtful, however. Although most AGN reside in rich environments, Seyfert galaxies, for instance, are found in any kind of environment just like normal galaxies (de Robertis *et al.*, 1998). The progenitors of AGN are thought to be starburst galaxies (Maiolino *et al.*, 2000), where a sufficient supply of gas is available for the formation of a black hole and accretion disk in the gravitational centre.

There are two main approaches to explaining the differences between AGN, in particular as far as jet power is concerned. The first possibility is that the jet power depends on the spin of the Kerr black hole (Wilson & Colbert, 1995). Processes such as the Blandford-Znajek mechanism (Sec. 2.2.3) are thought to become more efficient with increasing spin, which would result in more energy being funneled into the jets.

One argument speaking against this hypothesis is the fact that AGN with powerful jets (quasars) are observed predominantly at high redshifts, even though SBHs should be spun up with time by accretion of more matter. The second idea involves the environment in the centre of an active galaxy. It is intuitive that the amount of gas and dust available for accretion onto the SBH may determine the energy output in the form of relativistic jets. A theory involving this idea and based on the temporal evolution of AGN will be presented in the following.

AGN evolution

In order to explain the differences in AGN jet luminosities, an interesting evolutionary approach was proposed by Böttcher & Dermer (2002); Cavaliere & D'Elia (2002) and others. In this paradigm, giant elliptical galaxies are formed by mergers of smaller galaxies, mainly spirals, at cosmological redshifts $z > 2$ (probably significantly larger than 2). The merging process then results in star burst activity, and eventually AGN activity. These newborn AGN are thought to be IR-luminous QSO, *i.e. radio-quiet* AGN. This apparent paradox is thought to be caused by the fact that the very high accretion rates in early-type AGN, which are close to the Eddington limit and result from a high matter density around the central object, may give rise to optically thick accretion disks that extend all the way to the innermost stable orbit (see Sec. 2.2.2), quenching the formation of jets (see *e.g.* Böttcher & Dermer, 2002, and references therein). The maximum of the cosmological AGN formation rate is reached around $z \approx 2$.

As the supply of matter in the circumnuclear region is limited, the average matter density gradually decreases, and with it the accretion rate. At the same time, the black hole mass grows. Both decreasing accretion rate and increasing SBH mass result in a decreasing Eddington ratio (*i.e.* accretion disk luminosity), and an eventual emergence of radio jets. The jet formation may be the direct result of the optically thick accretion disk puffing up into an optically thin, advection dominated disk in the innermost part of the accretion flow (see also Sec. 2.2.2). The radio-loud AGN begin their existence as powerful quasars (or FR II radio galaxies when viewed edge-on), as there is still abundant matter to be accreted. As the supply of matter around the AGN is gradually depleted, the accretion rate decreases further; the accretion disk becomes less luminous and the jets begin to lose power. The AGN eventually turns into a BL Lac (FR I radio galaxy). This passive evolution appears to be quite stable and continues more slowly via LBL to the even less luminous HBL. The final stage of an evolving AGN is likely a normal, *i.e.* inactive giant elliptical galaxy, whose SBH has consumed most of the surrounding matter. This is in accord with the very low X-ray fluxes from giant ellipticals which are

known to harbour SBHs, consistent with very low accretion rates via ADAF disks. The latter are possibly remnants of previous nuclear activity.

The described evolutionary picture could explain why QSOs and quasars are only found at cosmological distances, and not in the local universe. BL Lacs on the other hand are expected at low redshifts, consistent with observations.[27] While all "normal" AGN seem to be associated with giant elliptical host galaxies – considerably larger than an average non-active elliptical galaxy – Seyfert nuclei are most often found in spiral galaxies. Since they are also found at low redshifts, are radio-quiet and have low luminosities, they do not fit into the evolutionary sequence just described. Presumably Seyferts are ordinary galaxies harbouring a SBH, much like our own galaxy, that were activated by an external event long after they have formed.

It is interesting to note that in the HST BL-Lac host galaxy survey (Urry *et al.*, 2000; Scarpa, Urry *et al.*, 2000), no fundamental morphological differences were found between normal massive elliptical galaxies and BL-Lac hosts, suggesting that the properties of the nucleus are more or less independent of the surrounding galaxy, and do not affect it. BL Lac hosts often appear to be dominating members of a mini-cluster of galaxies, supporting the merger theory.

2.4.3 Blazar sequence

The discovery of different types of blazars – FSRQ, LBL and HBL, which all exhibit the typical double-humped SED, but spanning different energy ranges – soon raised questions about whether these are just manifestations of the same type of object, but with some remaining slight differences in viewing angle, *i.e.* a de facto continuation of the unified AGN picture. This idea seemed to be supported by some initial studies, but later it was found that the supposed correlations between synchrotron peak frequency and viewing angle were most likely a result of observationally based selection effects. In another scenario, the "naked" AGN – which is essentially nothing more but a supermassive Kerr black hole – is placed into different physical environments, giving rise to different emission characteristics. The idea of the so-called *blazar sequence*, a unification of all three classes of blazars under the assumption that the differences between them could be explained by a continuous shift in some basic common parameters, was first introduced by Fossati, Maraschi *et al.* (1998). The underlying study was based on multi-frequency observations of a large sample of blazars, which led to the following observations:

[27] The fact that most observed BL Lacs are located at low redshifts is naturally only a selection effect due to more distant sources being too faint to be detected – which of course does not rule out the possibility that the co-moving density of BL Lacs is indeed greater at $z < 2$, as predicted.

1. The peak frequency of the high-energy component ($\nu_{\rm IC}$) correlates with that of the synchrotron component ($\nu_{\rm sync}$), so that their ratio $\nu_{\rm IC}/\nu_{\rm sync}$ is approximately constant.

2. The peak frequency of the synchrotron component is inversely correlated with the observed bolometric luminosity of the blazar.

3. The peak frequency of the synchrotron component is inversely correlated with the ratio of the luminosities of the high-energy peak and the synchrotron peak, $L_{\rm IC}/L_{\rm sync}$.

In this picture, the broad-band SED of a blazar is thus qualitatively determined by its synchrotron peak energy. For HBLs, which have the highest synchrotron peak frequencies (located in UV or X-rays), $L_{\rm IC}/L_{\rm sync} \approx 1$. As $\nu_{\rm sync}$ decreases, the luminosity ratio increases progressively to around 10 for LBLs and around 100 and above for FSRQ. The properties described above can be readily verified by comparing the SEDs of a number of blazars with a range of different values of $\nu_{\rm sync}$. A corresponding set of (averaged) SEDs is shown in Fig. 2.8. Note that the first property is a prediction of the SSC gamma-ray emission model described in Sec. 2.3.1.

The detailed prescriptions for the average blazar SEDs developed by Fossati, Maraschi *et al.* (1998) were later expanded by Donato *et al.* (2001), who among others deduced a constant ratio of $\nu_{\rm IC}/\nu_{\rm sync} \simeq 5 \cdot 10^8$ for most bolometric luminosities. This is of course a somewhat crude assumption, because it is tailored to the SSC model and thus does not take into account a possible hadronic contribution in the jet, nor does it depend on variables such as the energy of any EC target radiation fields present.

A theoretical framework for this observed trend, based on the SSC/EC model, was put forward by Ghisellini *et al.* (1998). The underlying idea is that a balance between particle acceleration and radiative cooling results in a reduced acceleration efficiency in high-power blazars: the higher the luminosity (for a given jet viewing angle), the higher the accretion rate and energy flow in the jet. The magnetic field strength inside the jet is hence thought to increase with blazar luminosity, and therefore also the resulting radiation losses of accelerated particles via synchrotron, and at higher energies also IC. This will effectively throttle the acceleration process and limit the maximum energy of the particles escaping the shocks. Magnetic field strengths fitted in SSC models are generally smaller than 0.1 G for HBL, in the order of 1 G for LBL and up to several G for FSRQ (Böttcher & Reimer, 2004), reflecting the increasing jet luminosity from one blazar type to the next. If the radiation density in the jet becomes very high (approaching that

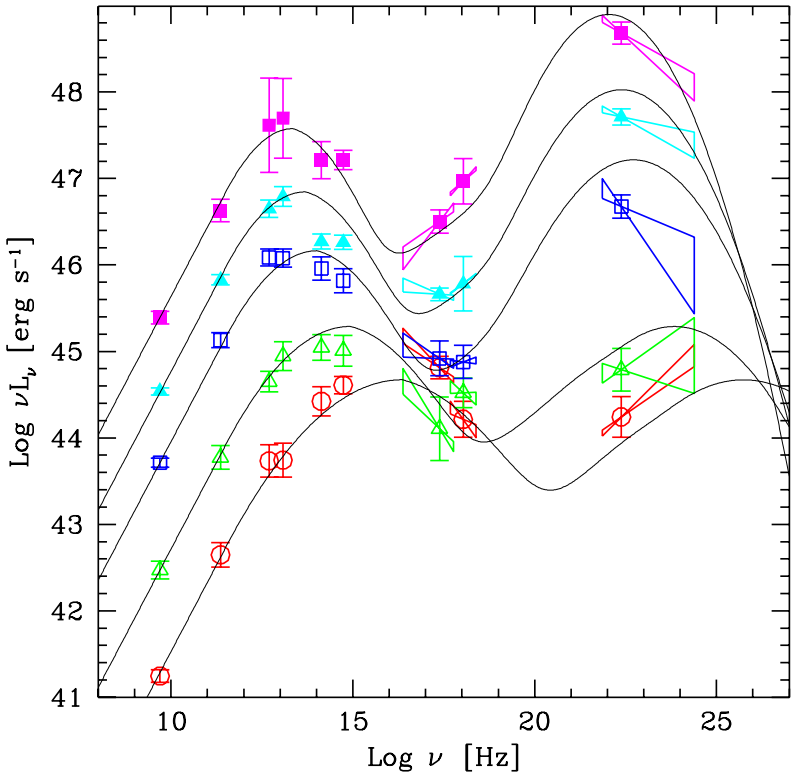

Figure 2.8: Averaged broad-band spectra of 40 blazars, divided into five groups according to their continuously distributed radio luminosities (which are equivalent to the synchrotron peak frequencies). The two spectra with the highest luminosities / lowest peak frequencies (*filled purple squares etc.*) are almost entirely made up of FSRQ, followed with decreasing luminosity by LBLs (*empty blue squares*) and finally HBLs (*empty red circles*). It can be seen that as the peak frequency increases, luminosity and high-energy to low-energy luminosity ratio decreases, defining the "blazar sequence". Taken from Donato *et al.* (2001).

required by hadronic emission models), the maximum energy of gamma rays that are able to escape from the jet may also be limited by pair production (see Sec. 2.3.2). In order to explain the increasing dominance of the IC-peak over the synchrotron peak when following the blazar sequence from HBLs via LBLs to FSRQ, increasing contributions from external photons (EC) need to be included in the modelling (Hartman *et al.*, 2001).[28] The physical explanation for this trend is a higher luminosity of the accretion disk and the jet itself in FSRQ when compared to HBLs, and correspondingly higher external photon fluxes injected into the jet. The luminosity in turn likely depends on the accretion rate, as explained in Sec. 2.4.2.

[28]Note that the ratio of IC-to-synchrotron flux is also sensitive to a change of SSC fit parameters, but only within certain limits.

Remarkable observational results

While all well established blazars do fit the blazar sequence picture, the X-ray detectors on board Swift (Gehrels *et al.*, 2004) have recently discovered an object which appears to contradict the predicted behaviour: ROXA J081009.9+384757.0, a distant (z = 3.95) FSRQ candidate, which seems to combine a very high luminosity ($> 10^{47}$ erg s^{-1}) with a hard X-ray spectrum, suggesting the synchrotron peak to be in the hard X-ray band (Giommi *et al.*, 2007). Ways to explain this observation include the postulation of a new blazar type present preferably at high redshifts, that the observed X-ray emission is not due to synchrotron radiation, or possibly an error in the redshift measurement.

Until recently, all blazars detected at VHE energies belonged to the HBL subclass. This is not surprising, as HBLs are thought to have their high-energy peak in the TeV energy range, whereas in LBLs and even more so FSRQ, the high-energy emission generally falls off to below the level of detectability already at energies well below 1 TeV. Furthermore, attenuation of gamma rays from the more distant FSRQ by the EBL is expected to be particularly severe. In 2005, however, MAGIC (Ferenc, 2005) discovered VHE gamma-ray emission above 80 GeV from the nearby (z = 0.069) LBL object *BL Lacertae* during a brief flare that lasted only for one observation night (Albert *et al.*, 2007). Despite the fact that the source remained below the detection limit for the rest of the observing campaign as well as during observations in the following year, this constitutes the first detection of VHE emission from a low-frequency peaked BL Lac object. As an even greater surprise came the detection of gamma rays above 110 GeV from the FSRQ 3C 279 by the same telescope one year later, again only during one observation night (Teshima *et al.*, 2007). 3C 279 is a rather well-studied blazar; it was the first such object detected by EGRET in MeV gamma rays and is known to be extremely variable. Its redshift is z = 0.536 and thus unprecedentedly large compared with that of all other detected VHE gamma-ray sources so far. The fact that VHE emission was still detected despite the large optical depth of intergalactic space to gamma rays for this redshift/energy combination (see the following section) gives us an idea of the extremely high energy output of this object. At the same time it also requires a very high Doppler factor and thus probably a very good alignment of the jet axis with the line of sight. As sensitivities of ground-based detectors to gamma rays in the GeV range are being further improved and space-borne detectors such as GLAST become operational, spectral information about the extreme high-energy end of the SEDs of other LBLs and FSRQ is likely to be forthcoming in the near future.

2.5 Extragalactic Background Light

One important physical process affecting our ability to observe gamma rays of energies above a few GeV from sources outside of our own galaxy is the attenuation of these gamma rays through collisions with photons of the diffuse *Extragalactic Background Light* (EBL), sometimes also referred to as *Diffuse Extragalactic Background RAdiation* (DEBRA). First proposed by Gould & Schréder (1967), the chief energy loss mechanism for VHE gamma rays is Bethe-Heitler pair-production with infrared-optical photons of the EBL:

$$\gamma_{\mathrm{VHE}} + \gamma_{\mathrm{EBL}} \longrightarrow e^+ + e^-$$

In the following we will focus on the various aspects of this process, an understanding of which is essential for interpreting observed gamma-ray spectra from extragalactic sources.

2.5.1 Spectral Energy Distribution of the EBL

The extragalactic background light, or *primordial* light, consists of the integrated electromagnetic radiation from all epochs, redshifted according to its formation epoch. In a slightly narrower definition which is commonly used and shall be adopted here in the context of gamma-ray absorption, the EBL spans the wavelengths from near UV all the way to far IR, bordering the Cosmic Microwave Background (CMB). Due to its large bandwidth, it is clear that measuring the spectrum of the EBL is far more difficult than, for instance, that of the CMB. Many different methods need to be employed to examine the various frequency ranges, and very different foregrounds need to be subtracted. In contrast to the CMB, which is thought to have been produced at a well defined time and reaches us from the current particle horizon, the EBL evolves together with the universe as it is continuously receiving contributions from stars, hot gas and dust in surrounding galaxies. It follows that its analysis and interpretation is intrinsically much more complex. Yet the importance of knowing the SED of the diffuse extragalactic background light to sufficient detail cannot be understated. While the Cosmic Microwave Background Radiation (CMBR) gives us a snapshot of the structure of the universe at the time of its production, the EBL provides us with essentially the same information for the time ever since the epoch of radiation-matter decoupling. It is dominated by (redshifted) thermal emission from stars: *directly* at optical and near infrared wavelengths, and *indirectly* through partial absorption of the latter and subsequent re-emission by dust at far-infrared wavelengths. As such it carries unique information about the galaxy

evolution and star formation history, *i.e.* about the structure formation of the universe since the decoupling of matter and radiation. Consequently, it depends strongly on the cosmological model and, if known to sufficient detail, could be used to constrain key cosmological parameters.

The main handicap of any physics related to the EBL is the fact that its spectrum is extremely difficult to measure, and hence poorly known. Direct measurements are hindered by strong foreground contamination from galactic dust and starlight, as well as interplanetary dust. Figure 2.9 shows the SED of the EBL according to current measurements, most of which are lower or upper limits; the large error bars reflect the difficulty of subtracting the various foreground emissions. The curves shown in the EBL plots are model predictions, which will be discussed in the following Sec. 2.5.2. The most striking features of the SED are the two broad peaks, one at near-infrared wavelengths ($0.8\mu\text{m} \lesssim \lambda \lesssim 3\mu\text{m}$) associated with red-shifted star light, and another in the far-infrared ($80\mu\text{m} \lesssim \lambda \lesssim 200\mu\text{m}$) resulting from absorption and re-emission of star light by interstellar dust. At large wavelengths, the EBL is joined by the CMBR, visible in the form of a sharp (blackbody) peak in the left panel of Fig. 2.9. The EBL in the mid-infrared wavelength region between 5 and 60 μm is particularly difficult to measure from Earth, due to a strong foreground from interplanetary dust emission. This is reflected in the virtual lack of direct measurements in the mid-IR range. For comprehensive reviews of the various techniques used to measure the EBL and a summary of existing models, the reader is referred to Hauser & Dwek (2001) and Kashlinsky (2005) for the periods up to 2001 and 2005, respectively.

2.5.2 EBL Models

On the theoretical side, efforts to predict the EBL are faced with serious challenges, because calculations require the complex treatment of key physical processes such as star and galaxy formation, most of which rely heavily on a number of cosmological assumptions (see *e.g.* Madau & Pozzetti, 1999; Primack *et al.*, 1999). Several different methods are used for modelling the EBL, which generally fall into one of three categories.

Empirical models

The *empirical* approach (see *e.g.* Malkan & Stecker, 1998, 2001; Stecker, Malkan & Scully, 2006) is based on observations, particularly of the number counts and mean spectral properties of galaxies as a function of redshift. Interpolation in redshift space (and extrapolation to higher redshifts) and integration over the number density then

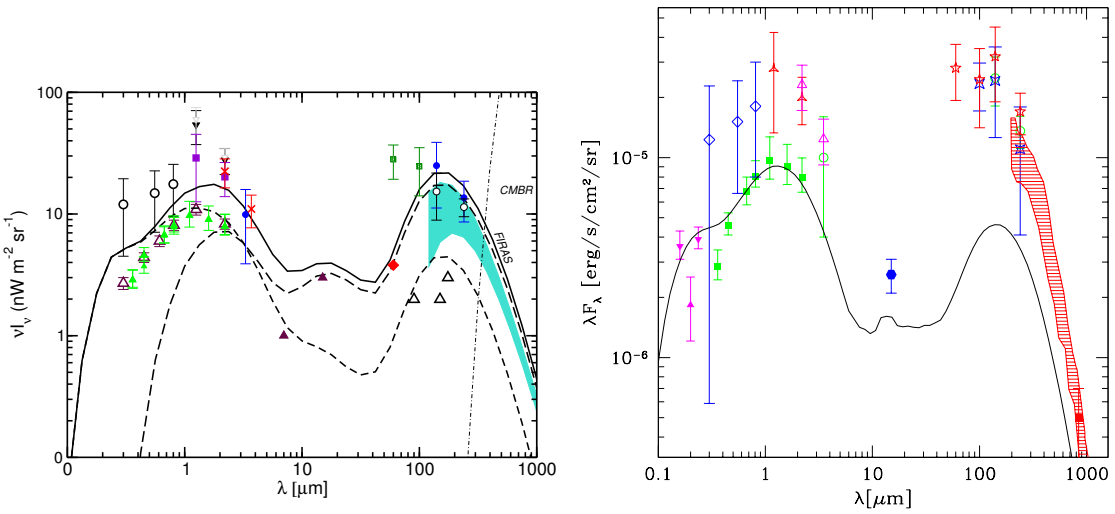

Figure 2.9: Spectral energy distribution of the Extragalactic Background Light (EBL) according to direct measurements (symbols on both plots), and two different models of the EBL. The quantities plotted on the y-axes are the energy fluxes per solid angle element. Two representative EBL models are shown. *Left*: Semi-empirical model of Kneiske, Mannheim & Hartmann (2002) (*solid line*), consisting of emission from optical galaxies (*short-dashed line*) and a component due to luminous infrared galaxies (*long-dashed line*). *Right*: Semi-analytic model of Primack *et al.* (2005) (*solid line*). The predicted fluxes of this model, at least in the optical and near-infrared, are likely on the low side, as they barely reach the lower limits due to integrated light from resolved sources of the HST Deep Field survey (*solid green squares*), and fall short of the lower limit from ISOCAM in the mid-IR (*filled blue hexagon*) and measurements of the FIRAS detector on board COBE in the far-IR (*shaded area*). For a complete explanation of measurements and models, see references.

produces the SED of the EBL at the present time ($z = 0$). In pure *backward evolution* (BE) models, instead of interpolating in redshift, only the (more easily observable) local galaxies are used as input, and galaxy contributions at higher redshifts are obtained by extrapolating via a galaxy evolution model. As it is based on actual data, the empirical approach requires the least number of assumptions, and compared to the other methods it is comparably simple. However, its obvious shortcoming is the fact that oversimplifications are unavoidable in regions of wavelength-redshift space for which measurements are unavailable or affected by large uncertainties.

Semi-analytic models

Much more complex, and hence much more dependent on theoretical assumptions, are the *semi-analytic* models (Primack *et al.*, 1999; Primack, Somerville *et al.*, 2001; Primack, 2002; Primack *et al.*, 2005). As suggested by the name, semi-analytic models, which are

forward evolution (FE) models, attempt to predict the evolution of galaxies and the intergalactic radiation field from first principles, starting with the very beginning of star formation and ending at the present epoch. In order to do this, rather detailed physical processes must be simulated, and a significant number of assumptions about the initial conditions must be made. Some of the most important input parameters are the star formation rate, the chemical evolution of stars, and the *initial mass function* (IMF), *i.e.* the distribution of masses of stars right after their formation. These models are based on a ΛCDM cosmology, and the chosen cosmological parameters also greatly affect the predicted EBL fluxes. While theoretically the most pure and potentially powerful approach, the predictions of semi-analytic models currently fail to fully agree with observations. Improvement can be achieved by adjusting certain parameters, but most often that is difficult to justify.

Hybrid models

Between these two extremes there exists a whole palette of intermediate approaches. For instance, there are FE models that use the same initial assumptions as semi-analytic models, but skip some of the detailed calculations by utilising observational results instead. The predictions of such models can be compared to the data, and initial conditions chosen so that the mismatch between the two is minimised. A hybrid between FE and purely empirical methods is the "semi-empirical" model of Kneiske, Mannheim & Hartmann (2002); Kneiske *et al.* (2004), which uses simulations to evolve stellar populations starting from well-defined initial conditions, while at lower redshifts it is closely matched to the data. A common problem of this latter type of models is that they do not take into account morphological evolution (including mergers) of galaxies and, probably as a consequence, suffer from a flux deficit in the far-IR part of the EBL spectrum. This problem is solved by an *ad hoc* introduction of a new class of galaxy, so-called *infrared galaxies*, which are thought to be galaxies surrounded by extended, star light absorbing and re-emitting dust clouds.

Evidently, all current models have their strengths and shortcomings. Particularly problematic for all types of models is the prediction of the SED of the "valley" in the mid-infrared part of the SED, because evolution models of the hot dust responsible for the emission in that part of the spectrum are currently very poorly constrained.[29] This part of the SED is also where currently the largest discrepancies between the predictions

[29]Ironically, as stated earlier, this is also the part of the spectrum concealed by re-emitted starlight from interplanetary dust, making a direct measurement very difficult.

of empirical and semi-analytic models are found: while the latter yield very low fluxes in the mid-IR ("deep valley"), the former contain observationally based emissions from warm dust and PAH molecules (Stecker & Scully, 2007), which have the effect of partially filling the mid-IR trough (see Fig. 2.9).

As far as gamma-ray astronomy is concerned, preferred EBL models are currently the semi-empirical models of Kneiske, Mannheim & Hartmann (2002) and Stecker, Malkan & Scully (2006), as well as the semi-analytic model of Primack *et al.* (Primack, Somerville *et al.*, 2001; Primack *et al.*, 2005) and renormalisations thereof. The solid lines in both plots of Fig. 2.9 correspond to the predictions of the first and third models.

2.5.3 Pair Production

Having introduced the intragalactic radiation field, the next question will naturally be how to quantify the attenuation experienced by VHE gamma rays travelling through this medium. Quantum electrodynamics (QED) provides us with the cross-section for pair production (Heitler, 1960),

$$\sigma_{\gamma\gamma}(E_\gamma, E_{\mathrm{EBL}}, \theta) \;=\; \frac{3\sigma_{\mathrm{T}}}{16}\,(1-\beta^2)\,\left[2\beta(\beta^2-2) + (3-\beta^4)\,\ln\frac{1+\beta}{1-\beta}\right] \tag{2.34}$$

where E_γ and E_{EBL} are the energies of the gamma ray and the background photon, respectively, measured in the lab frame of reference; θ is the angle between the directions of the two photons in the lab frame, and σ_{T} is the Thomson cross-section for electrons (see Eq. 2.26). The normalised velocity β of the produced electron and positron in the centre-of-mass system ($\beta \equiv v_e/c = \sqrt{1-1/\gamma^2}$) is:

$$\beta \;=\; \left(1 - \frac{2m_e^2c^4}{E_\gamma E_{\mathrm{EBL}}(1-\cos\theta)}\right)^{1/2} \tag{2.35}$$

At the peak of the pair-production cross-section, which at that point is maximised to $\sigma_{\gamma\gamma} \simeq 10^{-25}\mathrm{cm}^2$ (see *e.g.* Herterich, 1974), the energies of the gamma ray (E_γ) and the background photon (E_{EBL}) in the lab frame of reference, and under integration over angular distribution, satisfy the condition (Guy, Renault *et al.*, 2000)

$$E_\gamma \cdot E_{\mathrm{EBL}} \;\simeq\; (2\,m_e\,c^2)^2 \tag{2.36}$$

which is equivalent to

$$\lambda_{\mathrm{EBL}} \;\simeq\; 1.24\,\frac{E_\gamma}{1\,\mathrm{TeV}} \quad \mu\mathrm{m} \tag{2.37}$$

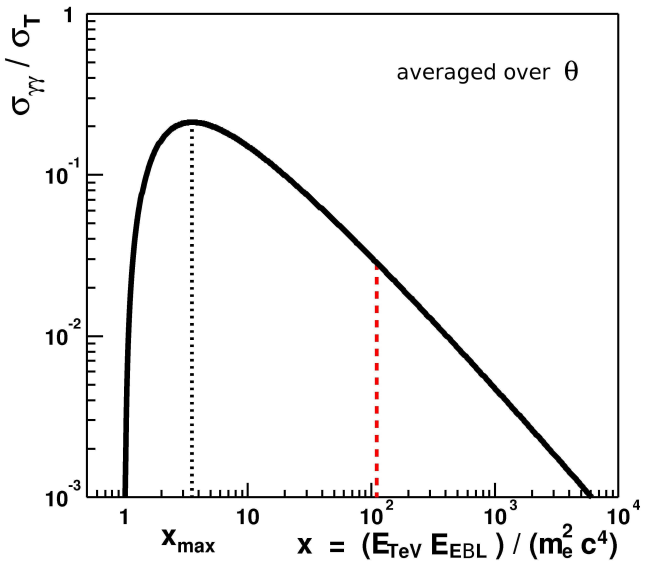

Figure 2.10: The pair-production cross-section according to Eq. 2.34, averaged over the angle between the two photons. The peak occurs at about $x_{\max} \simeq 3.6\, E_1 E_2 / m_e^2 c^4$. (For $\theta = 180°$(head-on collision), the peak would be at $x_{\max} \approx 2$.) At the red dashed line, the cross-section has fallen off to $1/e$ of its peak value. Adapted from Mazin (2003).

The cross-section as a function of the product of the two energies is plotted in Fig. 2.10. Fig. 2.11 shows the mapping between gamma-ray and background photon energy (in the local rest frame) for which the pair-production amplitude is maximised, in accordance with Eq. 2.36. Comparing this mapping with Fig. 2.9 reveals that the first EBL peak in the optical/near-IR is responsible for the strong increase of the attenuation above gamma-ray energies of $\sim 200\,\mathrm{GeV}$. We also conclude that the intergalactic medium is relatively transparent to gamma rays below $\sim 150\,\mathrm{GeV}$.

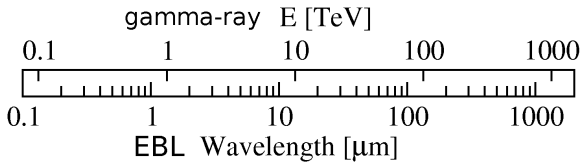

Figure 2.11: Correspondence of γ-ray to EBL energies for which $\sigma_{\gamma\gamma}$ is maximised.

2.5.4 Absorption of VHE gamma rays

From the magnitude of the absorption cross-section ($\sim \sigma_{\mathrm{T}}$) and typical EBL photon densities in the order of 0.1 cm^{-3}, it follows that the gamma-ray spectra of all extra-

galactic sources should be affected by absorption in the EBL. Naturally, one can expect the degree of attenuation to be proportional to the proper distance of the source from the observer, and thus roughly proportional to the redshift of the source. Furthermore, bearing in mind that the flux of background photons increases monotonically with decreasing energy (note that the differential EBL fluxes in Fig. 2.9 are plotted in units of νF_ν, not F_ν), the severity of the absorption will also increase rapidly with increasing gamma-ray energy.

The expected attenuation factor for a source at redshift z can be written in terms of the (strongly energy dependent) absorption coefficient $\chi(E, z)$ with respect to pair production with the EBL, which in this case is defined as

$$\frac{\mathrm{d}^2 N}{\mathrm{d}E_\gamma \mathrm{d}x} \;=\; -\chi(E_\gamma, z)\,\frac{\mathrm{d}N}{\mathrm{d}E_\gamma} \tag{2.38}$$

where $(\mathrm{d}N/\mathrm{d}E_\gamma)(x)$ is the differential gamma-ray flux at energy E_γ and x the space coordinate in the direction of propagation. Integration of Eq. 2.38 yields for the observed gamma-ray flux $\mathrm{d}N_{\mathrm{obs}}/\mathrm{d}E_\gamma$ at distance d from the source:

$$\frac{\mathrm{d}N_{\mathrm{obs}}}{\mathrm{d}E_\gamma}(E_\gamma, z) \;=\; e^{-\chi(E_\gamma, z)\, d}\,\frac{\mathrm{d}N}{\mathrm{d}E_\gamma}(E_\gamma) \tag{2.39}$$

where $\mathrm{d}N/\mathrm{d}E_\gamma$ is the intrinsic gamma-ray flux, *i.e.* the flux at the source ($d = 0$). The energy and redshift dependent mean free path $d_{\gamma\gamma}$ of VHE photons in the radiation field of the EBL is defined, as usual, to be the reciprocal of the absorption coefficient, *i.e.* $d_{\gamma\gamma}(E, z) = 1/\chi(E, z)$. Due to the sharp peak from direct starlight in the spectrum of the EBL at optical to near-IR wavelengths (Fig. 2.9) and the preferred absorber energy being inversely proportional to the gamma-ray energy (Eqns. 2.36 and 2.37), χ and $d_{\gamma\gamma}$ depend very strongly on energy. Therefore, when quantifying the gamma-ray absorption for a given redshift and energy, the two factors of the exponent in Eq. 2.39 are usually combined into a dimensionless *optical depth* $\tau_{\gamma\gamma}$:

$$\tau_{\gamma\gamma} \;\equiv\; \chi(E_\gamma, z)\, d(z) \;=\; \frac{d(z)}{d_{\gamma\gamma}(E_\gamma, z)} \tag{2.40}$$

With that definition, Eq. 2.39 becomes simply

$$\frac{\mathrm{d}N_{\mathrm{obs}}}{\mathrm{d}E_\gamma}(E_\gamma, z) \;=\; e^{-\tau_{\gamma\gamma}(E_\gamma, z)}\,\frac{\mathrm{d}N}{\mathrm{d}E_\gamma}(E_\gamma) \tag{2.41}$$

Thus, in order to predict the absorption of gamma rays from extragalactic sources by the EBL, the optical depth $\tau_{\gamma\gamma}$ as a function of energy and redshift needs to be known. To

first order, $\tau_{\gamma\gamma}(E_\gamma, z)$ is proportional to the source distance d and the average number density of background photons between source and observer, in a small window around the energy given by Eq. 2.36. As an example, for gamma-ray energies around 1 TeV, which pair-produce with EBL photons of the peak in the near-infrared ($\lambda_{\rm EBL} \approx 1.5\mu$m), the mean free path can be approximated by

$$d_{\gamma\gamma} \;\approx\; (0.26\,\sigma_{\rm T}\,n_{\rm IR})^{-1} \tag{2.42}$$

where $n_{\rm IR} \simeq 0.1$ cm^{-3} is the approximate intergalactic photon density in the near-infrared as suggested by contemporary EBL-models (Neronov & Semikoz, 2006). According to its definition (Eq. 2.40), the optical depth $\tau_{\gamma\gamma}$ is just the source distance d divided by $d_{\gamma\gamma}$.

While the above method is useful for making rough estimates of $\tau_{\gamma\gamma}$, a proper determination of the optical depth requires knowledge of the actual SED of the EBL. To this aim, it must first be remembered that energies as well as densities are affected by the cosmological redshift. This means that due to the expansion of space, the energies of the gamma rays and background photons we observe (*i.e.* at z = 0), $E_\gamma \equiv E_\gamma^0$ and $E_{\rm EBL} \equiv E_{\rm EBL}^0$, are reduced by a factor of $(1+z)^{-1}$ with respect to their energies at redshift z:

$$E_\gamma^z \;=\; (1+z)\,E_\gamma^0 \qquad \text{and} \qquad E_{\rm EBL}^z \;=\; (1+z)\,E_{\rm EBL}^0 \tag{2.43}$$

If $\mathrm{d}n(z,E)/\mathrm{d}E$ is the specific photon density as a function of redshift and energy, measured in units of [cm^{-3} erg^{-1}], and $F(E)$ the energy flux in units of [erg cm^{-2} s^{-1} srad^{-1}], then the connection between the (measured or predicted) isotropic EBL flux $F(E_{\rm EBL}^0)$ and the EBL photon density $\mathrm{d}n(0,E_{\rm EBL})/\mathrm{d}E_{\rm EBL}$, both at $z = 0$, is

$$\frac{\mathrm{d}n(0,E_{\rm EBL})}{\mathrm{d}E_{\rm EBL}} \;=\; \frac{4\pi}{c}\,\frac{F(E_{\rm EBL}^0)}{(E_{\rm EBL}^0)^2} \tag{2.44}$$

In the most general case of an evolving EBL model, the latter may be described by the specific number density of background photons as a function of energy and redshift $\mathrm{d}n(z,E_{\rm EBL})/\mathrm{d}E_{\rm EBL}$. The optical depth $\tau_{\gamma\gamma}(E_\gamma^0, z)$ for gamma rays with observed energy E_γ^0 from a source at redshift z can then be calculated by integrating over redshift z', the cosine of the angle θ between the two photons in the CM-system, and the background photon energy $E_{\rm EBL}$:

$$\tau_{\gamma\gamma}(E_\gamma^0, z) \;=\; \int_0^z \mathrm{d}z'\,\frac{\mathrm{d}l}{\mathrm{d}z'}\int_{-1}^{1}\mathrm{d}\cos\theta\,\frac{1-\cos\theta}{2}\int_{\epsilon(z')}^{\infty}\mathrm{d}E_{\rm EBL}\,\frac{\mathrm{d}n(z',E_{\rm EBL})}{\mathrm{d}E_{\rm EBL}}\,\sigma(E_\gamma^{z'}, E_{\rm EBL}, \theta) \tag{2.45}$$

where H_0 is the Hubble constant, $\sigma(E_\gamma^{z'}, E_{\rm EBL}, \theta)$ the cross-section for pair-production (Eq. 2.34) and $E_\gamma^{z'}$ the redshift-corrected energy of the gamma ray (Eq. 2.43). The lower limit for the integral (ϵ) is the energy threshold for the pair production, usually taken to be

$$\epsilon(z') \equiv E_{\rm EBL}^{thr}(z') = \frac{2m_e^2c^4}{E_\gamma^{z'}(1-\cos\theta)} \tag{2.46}$$

Finally, the relationship between redshift z and proper distance l is prescribed by the *proper line element*

$$\frac{{\rm d}l}{{\rm d}z} = \frac{c}{H_0} \frac{1}{(1+z)\sqrt{\Omega_M(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda}} \tag{2.47}$$

Above Eq. 2.47 is also known as *redshift-distance relation*. It depends on the cosmological parameters, which according to the recent results by WMAP are likely close to the following values: Hubble constant $H_0 = 71$ km s^{-1} Mpc^{-1}, (dark and baryonic) matter density $\Omega_{\rm M} = 0.27$, dark energy density $\Omega_\Lambda = 0.73$, and spatial curvature $\Omega_K \simeq 1 - \Omega_\Lambda - \Omega_M = 0$ (*i.e.* flat universe).

Many models of the EBL, such as those based mainly on observations, only predict the background photon density ${\rm d}n(z=0, E_{\rm EBL})/{\rm d}E_{\rm EBL}$ at present epoch, *i.e.* do not provide information about the evolution of the co-moving flux density with redshift. In that case the EBL spectrum is assumed to have remained unchanged over time, and only the cosmological density correction

$$\frac{{\rm d}n(z, E_{\rm EBL})}{{\rm d}E_{\rm EBL}} = (1+z)^3 \frac{{\rm d}n(0, E_{\rm EBL})}{{\rm d}E_{\rm EBL}} \tag{2.48}$$

is applied, and substituted in Eq. 2.45. Under these assumptions, the input photon densities at $z = 0$ can be directly obtained from measured fluxes via Eq. 2.44.

If the source redshift is very small ($z \ll 1$), the integral over redshift can be replaced by the Hubble distance $d_{\rm H} = cz/H_0$. Since in this case a redshift correction is unnecessary, Eq. 2.45 becomes

$$\tau_{\gamma\gamma}(E_\gamma, z) = \frac{cz}{H_0} \int_{-1}^{1} \frac{1-\cos\theta}{2} {\rm d}\cos\theta \int_\epsilon^\infty \frac{{\rm d}n(0, E_{\rm EBL})}{{\rm d}E_{\rm EBL}} \sigma(E_\gamma, E_{\rm EBL}, \theta)\, {\rm d}E_{\rm EBL} \tag{2.49}$$

Considering the large uncertainties in the EBL models and the small redshifts of detectable gamma-ray sources, the above expression (2.49) is often sufficiently accurate for calculating optical depths.

Various groups have calculated gamma-ray attenuation curves ($\tau_{\gamma\gamma}$ as a function of energy and redshift) from their EBL models (see *e.g.* Malkan & Stecker, 2001; Primack,

Somerville *et al.*, 2001; Kneiske *et al.*, 2004; Primack *et al.*, 2005). Due to the currently poorly constrained EBL fluxes, reflected by the discrepancies between the different models, the results are of course subject to large systematic errors. Figure 2.12 (left plot) shows the transmission coefficient $e^{-\tau_{\gamma\gamma}}$ as a function of redshift and energy as calculated by Primack *et al.* (2005), illustrating the strong energy dependence. The rapid steepening of the transmission curves at energies above about 200 GeV is due to the absorbing photon's energy dropping to optical/near-infrared wavelengths, where the EBL density exhibits a sharp peak (see Fig. 2.9). The flattening between 1 and ∼4 TeV is due to the trough in the EBL spectrum at mid-infrared wavelengths, and the following second drop-off due to the peak in the far-infrared. A different way of plotting the attenuation as a function of gamma-ray energy is shown in the right panel of Fig. 2.12; for the same model, a number of so-called *attenuation edges*, which are the redshifts at which the absorption coefficient has a constant value, are plotted as a function of energy. This way of depicting the attenuation has its usefulness in the fact that one can read off a direct estimate of how far gamma-ray telescopes can "see" as a function of their energy thresholds. Note that the particular model used to obtain these plots (that of Primack *et al.*, 2005) represents the very low end of the EBL flux range currently considered realistic.

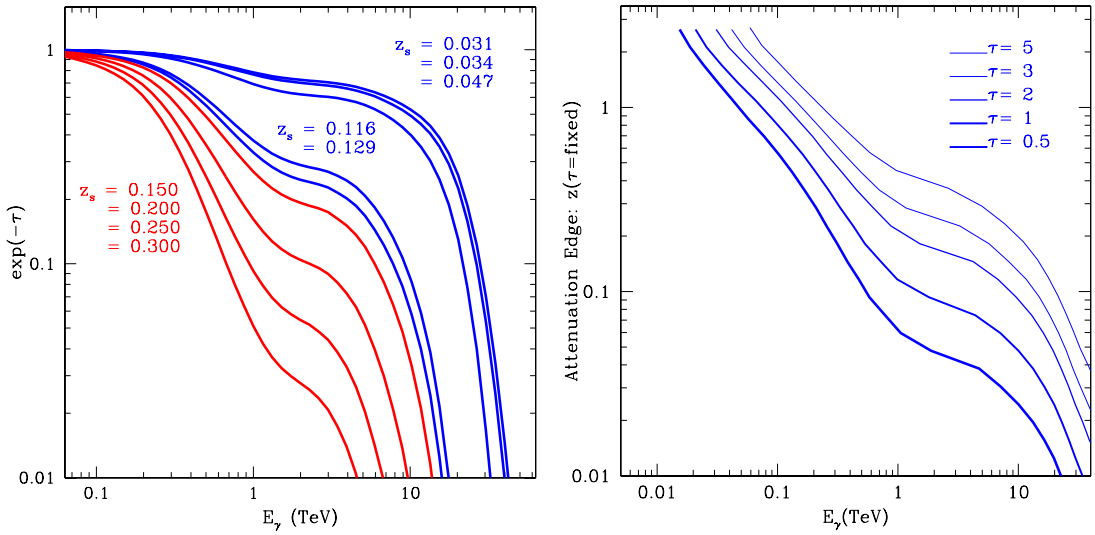

Figure 2.12: Dependence of the optical depth for gamma-ray absorption in the EBL on gamma-ray energy and source redshift. *Left*: Transmission coefficient $e^{-\tau}$ (here $\tau \equiv \tau_{\gamma\gamma}$ is the optical depth) for pair-production of gamma rays in the EBL as a function of energy, plotted for various values of the source redshift $z_{\rm s}$. *Right*: Redshift z for which the transmission coefficient has the value $\mathrm{e}^{-\tau}$, for five different values of the optical depth τ, as a function of gamma-ray energy. These particular attenuation curves are derived from the EBL model shown in the right panel of Fig. 2.9 (Primack *et al.*, 2005).

In order to facilitate the use of EBL absorption factors, empirical functions can be fitted to the absorption curves; *e.g.* using the EBL model of Primack, Somerville *et al.* (2001), Stecker & Scully (2006) derived a redshift-dependent analytic logarithmic approximation of the absorption coefficient for the energy range from 0.2 to 2 TeV.

Using the energy and redshift dependent optical depths calculated from any given EBL model, one can readily correct the observed VHE gamma-ray spectra of blazars for the absorption effect to recover the "de-absorbed", *i.e.* intrinsic blazar spectra. The apparent straightforwardness[30] of this step is however compromised by the large uncertainties in current EBL models, which are of orders of magnitude even in the most relevant near-infrared band. Disentangling the effects of intrinsic blazar spectra and EBL spectrum therefore yields reliable results currently only in a limited region of the energy-redshift parameter space, or more specifically, at low redshifts and/or low gamma-ray energies. Correcting the observed spectra for EBL absorption is of course essential for any meaningful physical interpretation of the emission spectra. Depending on the employed EBL model and the observed source, the results are more or less conclusive (for examples, see *e.g.* Konopelko *et al.*, 2003; Costamante *et al.*, 2003, 2004; Kneiske *et al.*, 2004; Konopelko, Mastichiadis & Stecker, 2005; Costamante, 2007).

In the hope that a particular chosen EBL model is reasonably realistic, the derived intrinsic VHE spectra can then be fitted with gamma-ray emission models. For many of the detected VHE-emitting blazars, this two-step procedure seems to yield sensible results; in particular the combination of more recent, low-flux (see below) EBL models and the SSC model seem to describe all HBLs closer than z = 0.129 (the redshift of H 1426+428) very well. Whenever higher redshifts are involved, however, current blazar modelling seems to be hitting an obstacle. As one can see from the representative case of H 1426+428 (see Sec. 2.6.5), slightly different EBL parameterizations lead to very different emission characteristics – giving rise to a degeneracy that is difficult to resolve.

While the study of gamma-ray emission mechanisms in extragalactic sources is greatly complicated by the absorption of VHE gamma rays in the extragalactic radiation field, the fact that this interaction occurs provides us at the same time with a unique tool, independent of direct measurements, to probe the SED of the EBL at optical to far-infrared wavelengths. Not long after demonstrating that realistic emission spectra could be recovered by correcting observed gamma-ray data for absorption, the opposite approach was taken, namely to use EBL-absorption factors derived from VHE gamma-ray

[30] Strictly speaking this is not quite true, for even if the spectrum of the EBL were known precisely, some ambiguities arise from the fact that several different EBL shapes can leave the same "imprint" on a given blazar spectrum (Mazin & Raue, 2007).

observations of blazars to constrain normalisation and shape of the EBL-SED (see, for instance, Mapelli, Salvaterra & Ferrara, 2004; Schroedter, 2005; Mazin & Raue, 2007). This can be done as long as sufficient knowledge about the intrinsic spectra of the emitting objects is available. As one may suspect from what was said in Sec. 2.3 however, the latter requirement is only marginally fulfilled to date, leaving plenty of room for spectral modifications due to processes that are not accounted for in the particular models used. One can only hope to determine intrinsic blazar spectra observationally by "de-absorbing" the measured spectra with a well known EBL flux, or vice versa, constrain the EBL flux at the relevant energies by applying a known blazar SED. Unfortunately, neither blazar nor EBL spectra are well known to date.

In spite of these caveats, upper limits on the EBL flux were derived by making some "reasonable" assumptions about the intrinsic blazar spectra. These assumptions are normally based on

- existing gamma-ray emission models, in particular the SSC/EC model
- the measured VHE spectra of low-redshift BL Lacs such as Mrk 421 and Mrk 501, which are only weakly affected by EBL absorption at the highest energies

For instance, Aharonian *et al.* (2006b) and Costamante (2007) used the VHE spectra of two distant HBLs recently detected by HESS, 1ES 1101-232 and H 2356-309, to derive very stringent upper limits on the EBL for the region of the starlight peak in the optical/near-IR bands. These limits are valid under the requirement that the intrinsic spectral indices of these blazars are no harder than $\alpha = 1.5$, an assumption which one can attempt to justify from the observed VHE spectra of the two nearby Markarian blazars, and common SSC model fits of the SEDs of other HBL. The obtained limits seem to suggest that the actual optical to near-IR EBL fluxes are only about a factor of two above the lower limits from galaxy counts, and thus considerably lower than assumed in previous EBL models.[31] Whether this method of constraining the EBL is viable or not is still heavily disputed, as it stands and falls with the assumption that $\alpha \leq 1.5$. For example, Katarzyński *et al.* (2006) were able to fit the broad-band spectra of several BL Lacs with single-zone SSC models that give rise to hard intrinsic VHE spectra with spectral indices up to $\alpha \geq 2/3$, and Stecker, Baring & Summerlin (2007) (see also Stecker & Scully, 2007) used relativistic shock acceleration simulations to argue

[31] Apart from the already mentioned problem arising from the poor knowledge of the blazar emission spectra in the VHE range, one could argue this particular work is slightly biased in that a given EBL-SED shape (that of Primack *et al.*, 1999) is assumed (although not without a physical justification). This shape is simply scaled until the derived intrinsic spectral index matches a maximum permissible value of $\alpha = 1.5$, instead of allowing all parameters to vary.

that $\alpha \gtrsim 1$ should be possible. Such hard spectra in turn would allow for EBL densities much higher than the limits derived in above studies.

Somewhat more conservative (although therefore also less constraining) are approaches such as that of Dwek & Krennrich (2005) (see also Dwek, Krennrich & Arendt, 2005), who used a series of EBL templates to derive the intrinsic SEDs of Mrk 421, Mrk 501 and H 1426+428, and were able to rule out many EBL spectra that led to unphysical behaviour of the blazar VHE spectra, such as a spectral hardening with energy. More recently, Mazin & Raue (2007) made use of all available measurements of VHE spectra of blazars to derive robust upper limits on the EBL flux all the way from optical to far-infrared wavelengths.

Under the assumption that our knowledge about the actual emission spectra of HBLs (or more generally, blazars) will be improved in the future, these objects are likely to become reasonably accurate "test beams" for measuring the EBL spectrum, particularly when observed at several frequencies simultaneously during multiwavelength campaigns. The improved EBL models in turn could then be used to unravel the intrinsic spectra of other, less well-known extragalactic gamma-ray sources.

2.6 H 1426+428

We shall now examine one of the confirmed VHE gamma-ray emitting BL Lac objects, H 1426+428, in more detail, and elaborate why the study of this particular blazar is of special interest to high-energy astrophysics as well as cosmology. Past multiwavelength observations of H 1426+428 revealed several quite untypical properties of this object. Thus, observations with the new generation of gamma-ray telescopes, with improved sensitivities that extend well into the poorly explored energy region between $\sim$1 and 300 GeV, promise to provide us with valuable information about the details of gamma-ray production mechanisms at work in high-energy peaked BL Lacs, capable of accelerating charged particles to the highest energies observed to date. If the extreme properties of H 1426+428 are confirmed, existing gamma-ray emission models (and more generally, blazar models) might need to be expanded, or alternative emission models considered. On the other hand, the source is situated at an optimum distance for probing the extragalactic background light, whose spectrum is presently very poorly constrained. As we will see in the following, the intrinsic properties of this object paired with its relatively large distance makes H 1426+428 a unique tool for testing existing models like probably no other blazar observed to date.

2.6.1 Discovery and background

The extragalactic object named H 1426+428, often also referred to as 1ES 1426+428, is located in the constellation of Bootes at an RA of 14h 28' 32.60" and a declination of +42° 40' 29.0" in J2000 coordinates (see star chart in Fig. 2.13). The elliptical ($\epsilon \simeq 0.33$; Falomo, Scarpa, Treves & Urry, 2000) host galaxy of H 1426+428 has an apparent visual magnitude of 16 and an approximate half-light[32] diameter of 4.50 arcseconds, as measured by the Hubble Space Telescope (HST; Urry *et al.*, 2000; Scarpa, Urry *et al.*, 2000). Its cosmological redshift was measured to be z = 0.129 (see Sec. 2.6.2), which corresponds to a distance d of about 460 Mpc when using the correct redshift-distance relation with common cosmological parameters (see Eq. 2.47 and text below). Using the same z and cosmological assumptions, apparent magnitude and size translate into an absolute visual magnitude $M_{\rm V} = -23.7$, which is close to average for BL Lac host galaxies, and a half-light diameter of 14.1 kpc. The total size of the elliptical galaxy was estimated by applying a *de Vaucouleurs* law fit to the light distribution, yielding a diameter of ∼44 kpc.[33] An optical contour plot of H 1426+428 's elliptical host galaxy can be seen in Fig. 2.14. The mass of the galaxy's central black hole is estimated to be $5.3^{+2.3}_{-1.9} \cdot 10^8$ solar masses (Wu *et al.*, 2002), a very typical value for blazars.

H 1426+428 is an *X-ray selected* BL Lac, which is to say that it was discovered through optical spectroscopic identification of a point-source discovered in X-ray surveys. First evidence for X-ray emission from the direction of H 1426+428 was provided in the early 1970's by the SAS-1 satellite (Giacconi *et al.*, 1972); this X-ray source was henceforth catalogued as *4U 1444+43*.[34] More than 10 years later, the source was "rediscovered" by the HEAO-I X-ray survey in the 2 - 6 keV band (Wood *et al.*, 1984), assigning the object its current name. Eventually, Remillard *et al.* (1989) discovered an optical counterpart of the X-ray source, an elliptical galaxy whose redshift the group determined to be 0.129. They also observed the light to have a high degree of linear polarisation, which together with the absence of emission and absorption line features in the spectrum identified it as a BL Lac object. This was later confirmed by observations with *XMM-Newton* (Blustin, Page & Branduardi-Raymont, 2004). Recently, radio observations with the VLBI array revealed a parsec-scale jet in H 1426+428, exhibit-

[32] In the literature, "half-light radius" refers to the distance from the centre of the galaxy at which the light intensity falls off to half its peak value.

[33] For comparison, the Milky Way, a conventional spiral (disk) galaxy, has $M_{\rm V} = -20.7$ and a diameter of about 27.5 kpc.

[34] The "U" in the source name stands for *Uhuru*, which means "freedom" in Swahili and is the more popular name of the Kenya-launched SAS-1 satellite.

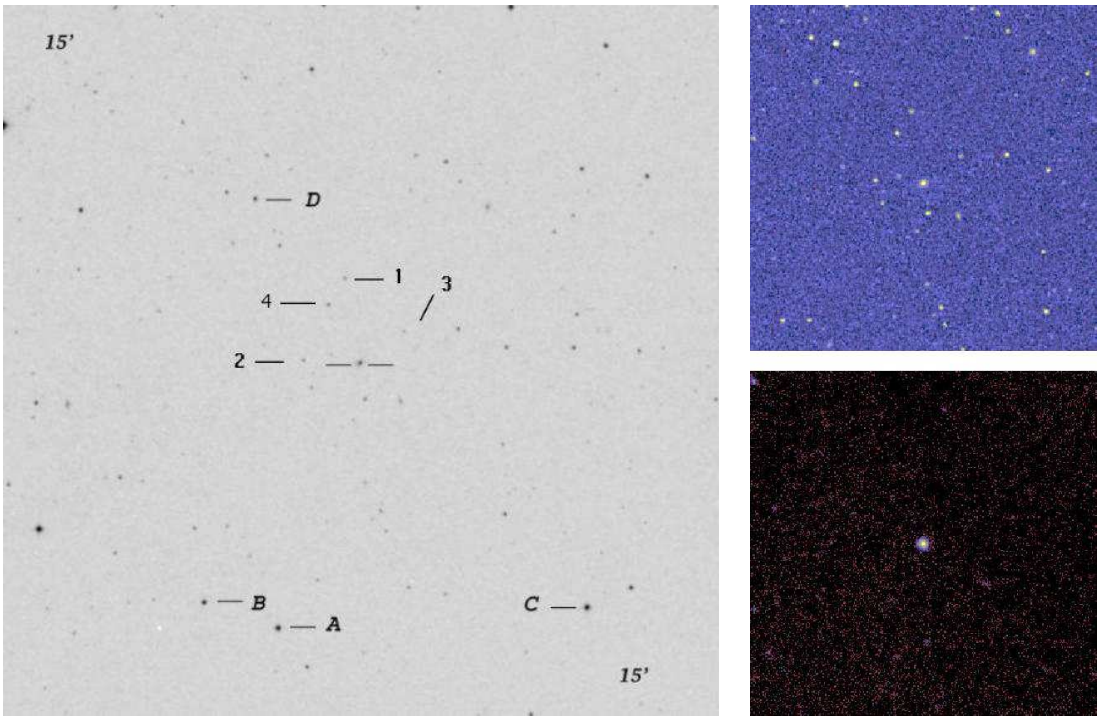

Figure 2.13: *Left*: Sky chart of a 15×15 arcmin square area around H 1426+428 (centered), taken from Smith *et al.* (1991). The star labelled "*C*" has a V-magnitude $M_{\mathrm{V}} = 13.46$. *Right*: Images of the sky region centered on H 1426+428 at visible (*top*) and X-ray (*bottom*) wavelengths (from www-ROSAT, 2006).

ing superluminal velocity with an apparent speed of $(2.09 \pm 0.53)\, c$; see Piner, Pant & Edwards (2008).

2.6.2 Gamma-ray observations of H 1426+428

As we will see in Sec. 2.6.3, the X-ray radiation from H 1426+428 is thought to be entirely due to synchrotron emission from its relativistic jet. In accordance with all current gamma-ray emission models (see Sec. 2.3), the accelerated charged particles should give rise to an emission of highly energetic radiation peaking at TeV energies. The source was thus labelled as a candidate for VHE gamma-ray emission and a promising target for ground-based gamma-ray telescopes (Costamante *et al.*, 2001). When the CGRO (see Sec. 1.5.2) was launched in the 1990's, EGRET surveyed the whole sky for gamma rays between 20 MeV and ~ 1 GeV, but did not detect any emission from the direction of H 1426+428. In the very late 1990's, ground-based atmospheric Cherenkov detectors had become sensitive enough to be able to resolve gamma-ray emitters fainter than the Crab Nebula, and observation campaigns by CAT (1998, 1999 and 2000; Djannati-Ataï

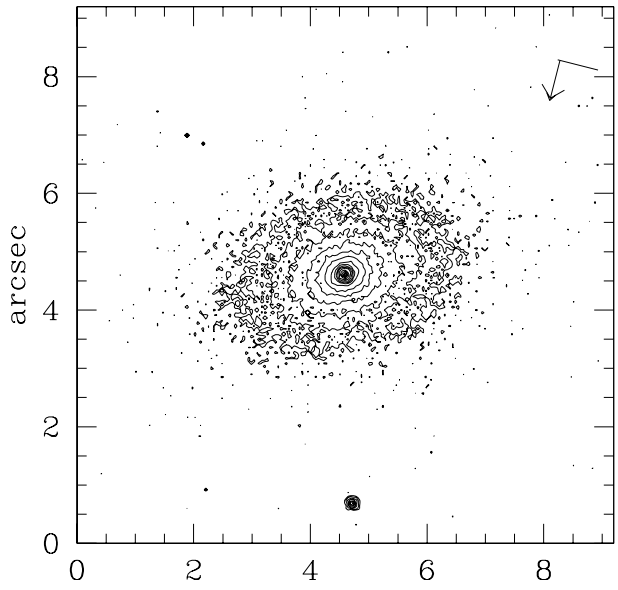

Figure 2.14: Optical contour plot of H 1426+428's host galaxy, from an image taken 1996 with the Hubble Space Telescope (HST). Celestial North is indicated by the arrow in the upper right corner. Credits to Scarpa, Urry *et al.* (2000).

et al., 2002), HEGRA (1999 and 2000; Aharonian *et al.*, 2002) and Whipple (1995-2001; Horan *et al.*, 2002) all eventually revealed a gamma-ray signal above 250 GeV, firmly establishing H 1426+428 as an emitter of TeV gamma rays.[35]

All three groups were able to extract spectral information from their data: CAT from 250 GeV to 1 TeV (Djannati-Ataï *et al.*, 2002), Whipple from 369 GeV to 1.71 TeV (Petry *et al.*, 2002) and HEGRA from 780 GeV all the way up to 10.2 TeV (Aharonian *et al.*, 2003). The spectral results of Whipple and CAT at the lower energies are compatible with a simple power law of index $\alpha = 3.5$ (in detail, the best-fit spectral indices are $\alpha = 3.66 \pm 0.41$ for CAT and $\alpha = 3.50 \pm 0.35$ for Whipple), whereas the HEGRA spectrum appears to flatten above ~ 1 TeV. Such a signature, *i.e.* a steep spectrum at GeV energies with a spectral hardening occurring in the TeV range, is expected from sources distant enough to be measurably affected by absorption of the original gamma-rays by the EBL (see Sec. 2.5), an effect becoming more pronounced with increasing blazar distance. Indeed, the VHE gamma-ray spectra of all detected blazars with redshifts below that

[35] In gamma-ray astronomy, a signal of at least 5 standard deviations above background (5 σ) is usually required in order for a source to be called "detected". Although 5σ may appear conservatively high, this is partially motivated by the need to safeguard against possible large systematic uncertainties affecting the technique of ground-based gamma-ray detection.

of H 1426+428 appear to obey simple power laws, and their spectra are harder. At a redshift of 0.129, H 1426+428 is at an intermediate distance when compared to the sample of all objects currently confirmed to emit gamma rays in the GeV-TeV range.[36] Most of the further BL Lacs exhibit spectral shapes in the TeV range that are quite similar to that of H 1426+428, suggesting that the observed spectral flattening above ∼1 TeV is indeed an imprint of the EBL rather than a deviation from a power law of the intrinsic spectrum. In Sec. 2.6.6 we will address the effects of gamma-ray absorption on the spectrum of H 1426+428 once more.

HEGRA performed extensive follow-up observations of H 1426+428 in 2002,[37] obtaining a spectral shape identical to that measured two years earlier, but observed a slight overall hardening of the spectrum (albeit not statistically significant). Also, more importantly, the observed gamma-ray flux above 1 TeV had decreased from about 8% of the Crab Nebula flux in 1999 and 2000 to ∼ 3% in 2002 (Horns *et al.*, 2003).[38] The left panel in Fig. 2.15 shows the two-dimensional significance map resulting from HEGRA's 1999-2002 stereo observations of the source. In 2004, the blazar was also observed by the CELESTE non-imaging atmospheric Cherenkov telescope, but no gamma-ray emission significant enough to be detected by that experiment was found above 80 GeV (Smith *et al.*, 2006). The resulting $2\,\sigma$ upper limit on the flux is $\nu F_\nu < 0.2 \cdot 10^{10}$ erg cm^{-2} s^{-1} at 80 GeV.

Finally, the results of the UHE gamma-ray observations of H 1426+428 with the HEGRA surface array should be mentioned. The latter consisted of an array of 224 scintillation counters (energy threshold ∼ 40 TeV) combined with a 7×7 array of open Cherenkov counters (AIROBICC; threshold ∼ 20 TeV). In addition, 17 Geiger towers were distributed across the detector area, intended to improve gamma-hadron separation. AIROBICC has provided upper flux limits in the order of 10^{-13} cm^{-2} s^{-1} for a number

[36] As the new generation of imaging ACT observe more and more sources, the gamma-ray horizon continues to move out to larger redshifts. For completeness, the seven furthest VHE gamma-ray emitters to date are the FSRQ 3C 279 (z = 0.536; Teshima *et al.*, 2007) and the HBLs 1ES 1011+496 (z = 0.212; Albert *et al.*, 2007b), 1ES 0347-121 (z = 0.188 Aharonian *et al.*, 2007), 1ES 1101-232 (z = 0.186 Aharonian *et al.*, 2007c), 1ES 1218+30.4 (z = 0.182; Albert *et al.*, 2006), H 2356-309 (z = 0.165; Aharonian *et al.*, 2006c) and 1ES 0229+200 (z = 0.139; Aharonian *et al.*, 2007b). VHE gamma-rays from PG 1553+113, a BL Lac with yet unknown redshift, were recently also discovered by both MAGIC and HESS. Using the unprecedentedly steep ($\alpha \simeq 4.1$) VHE-spectrum, Mazin & Goebel (2006) derived an upper limit of z < 0.69 for the redshift of that source.

[37] In fact, the main motivation for extending the lifetime of the HEGRA array of ACT by one year was to be able to take more data on this interesting BL Lac as part of a multi-wavelength campaign. Originally the detector's shutdown had been scheduled for 2001. HEGRA observed H 1426+428 for 217.5 hours in 2002.

[38] Even between the 1999 and 2000 observing periods, the flux detected by HEGRA seems to have decreased slightly (Götting, 2006).

of VHE gamma-ray emitters, including Mrk 421 and Mrk 501 (Mannheim, Westerhoff *et al.*, 1996). As part of a sky survey in the years 1994/1995, both detector components registered a significant and steady excess of events from the direction of H 1426+428, with an angular distribution compatible with that of a point source (Prahl, 1999). The statistical significance of the excess was 4.1 σ above background, and therefore just below the limit of firm detection. The right panel of Fig. 2.15 shows the resulting significance map. The energy threshold for these observations was 20.9 TeV, and the "detected" integral flux above threshold was (5.7 ± 1.3) cm^{-2} s^{-1}, corresponding to 4.5 times the Crab flux above the same threshold. The HEGRA ACT data were subsequently searched for excess events above 20.9 TeV, and the derived upper flux limit for the 1999/2000 data lies only a factor of 1.7 below the speculative flux measured by the surface array. For the 2002 data it is much lower (Götting, 2006).

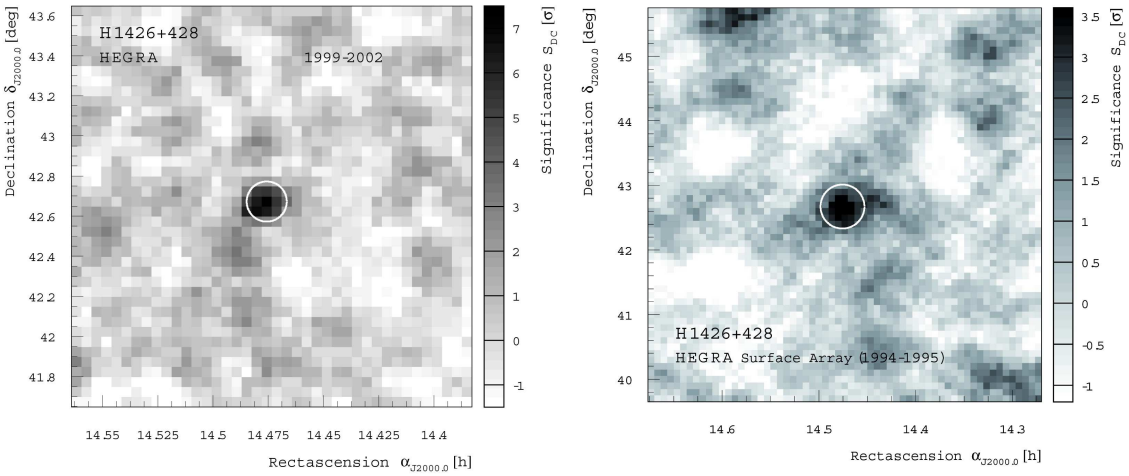

Figure 2.15: *Left*: Accumulated significance map for the 1999-2002 observations of H 1426+428 with the HEGRA array of imaging ACT at an energy threshold of 1.3 TeV. The radius of the white circle centered on the object is 0.1° and corresponds to the HEGRA angular resolution, the pixel size is 0.05°. *Right*: Significance map for observations with the HEGRA surface detector array (see text) with an energy threshold of 20.9 TeV. The circle denotes the angular resolution of the AIROBICC detector, which is 0.34°. Images were taken from Götting (2006).

While according to current models, such a high flux value above 20 TeV energy appears to be extremely unlikely (if only due to the expected high attenuation of gamma rays of such high energies), the likelihood of the observed excess being a statistical fluctuation is also low. The only way such a high flux above 20 TeV could be produced by H 1426+428 were if during the time of observation, the source's high-energy peak were indeed above 10 TeV, and the EBL flux in the more or less unconstrained mid-IR

region were lower than expected. It also appears intriguing that the VHE gamma-ray flux from this blazar seems to have been decreasing continuously between 1999 and 2002 (and possibly again from 2003 to 2004, see Sec. 7.4), which could be indicative of a long-term trend. In this case, one might consider the possibility of even higher VHE gamma-ray fluxes from that source before it was first observed by ACTs, such as at the time of the AIROBICC observations.

Our current knowledge of the VHE gamma-ray spectrum of H 1426+428 is summarised in Fig. 2.16, which includes the results of all published gamma-ray observations from 1998 to 2000, since for this period the flux level appeared to be relatively constant. The (lower flux) HEGRA results from 2002 are not included. Also shown are absorption-corrected fluxes for a high-flux EBL model with three slightly different normalisations, assuming that the intrinsic spectrum follows a simple power law within the covered energy range. It should be noted that all ground-based gamma-ray observations of H 1426+428 were made between the months of February and June, when the source is most easily visible in the night sky.

The redshift of H 1426+428

Since disentangling the effects of EBL absorption and the intrinsic emission spectrum of a gamma-ray source depend very sensitively on the distance of the object, knowing the exact redshift of H 1426+428 is of great importance. There have been two redshift measurements of the blazar's host galaxy to date. The original measurement was provided by Remillard *et al.* (1989), however the fact that their spectral acceptance did not cover the most prominent absorption feature for blazars, the Ca-II "H&K" lines, motivated D. Engels *et al.* (Götting, 2006) to verify the redshift through a new measurement. In September 2003, using the much more precise CAFOS spectrograph at the 2.2 m telescope of Calar Alto (Almería, Spain), Engels measured the positions of all available absorption lines and concluded that the blazar's redshift was indeed $z = 0.129 \pm 0.001$. The final spectrum is shown in Fig. 2.17.

2.6.3 X-ray observations of H 1426+428

Like most BL Lac objects, H 1426+428 is a strong X-ray emitter; its X-ray luminosity in the 2 - 6 keV band is in the order of 10^{44} erg s^{-1}. H 1426+428 has been classified as a high-energy peak BL-Lac type blazar (HBL) owing to the fact that its synchrotron emission peak ($\nu_{\rm peak}$ in νF_ν representation, corresponding to $E_{\rm peak} = h\nu_{\rm peak}$) is located in the X-ray regime. The synchrotron spectrum of H 1426+428 was observed to be highly

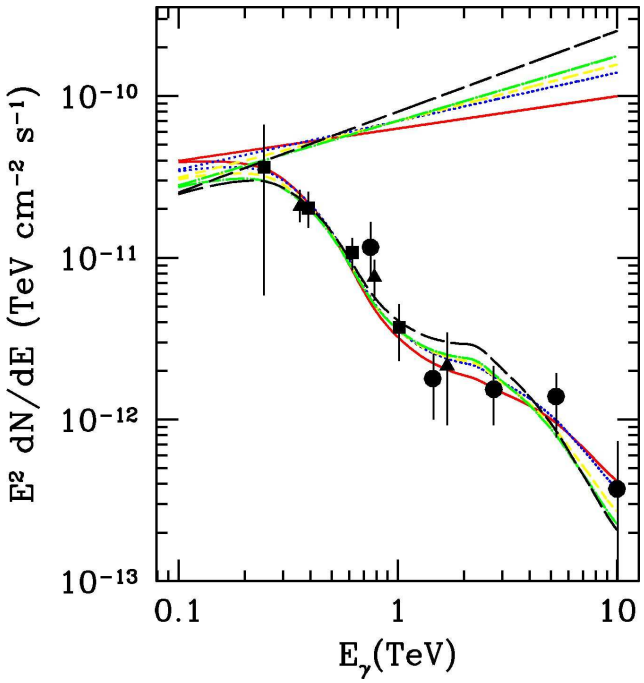

Figure 2.16: VHE gamma-ray spectrum of H 1426+428, including all published flux measurements from 1998-2000 (*squares*: CAT 1998-2000; *triangles*: Whipple 2001; *circles*: HEGRA 1999-2000); see text for full references. Also shown are fits to the observed fluxes under the requirement that the EBL-corrected fluxes follow a simple power law. This is done for three different EBL-parameterizations, which differ only by a normalisation factor for the mid- to far-IR background, while the optical and near-IR background is held constant. The normalisations are, from hardest (*green*) to softest (*red*) intrinsic spectrum: 3.8, 2.7 and 2.0 nW m^{-2} s^{-1}. Adapted from Mapelli, Salvaterra & Ferrara (2004).

variable in overall flux normalisation as well as energy scale. In February 1999, the LECS, MECS and PDS X-ray detectors on board BeppoSAX measured a spectrum consistent with a simple power law flatter than unity in the energy range between $\sim$5 and 100 keV, inferring a synchrotron peak energy of 100 keV or above (Costamante *et al.*, 2001).[39] From March to June 2002, the source was observed with the PCA spectrometer on board the *Rossi X-ray Timing Explorer* (RXTE), which has adequate sensitivity between 3 and 24 keV, as part of a multi-wavelength campaign. It was found that during the three-month interval, $E_{\rm peak}$ appeared to fluctuate significantly, varying from somewhere within the 3 - 24 keV interval to above 100 keV (Falcone *et al.*, 2004). Recently, H 1426+428 was

[39]The authors note that H 1426+428 was in a state of low activity during the BeppoSAX observations, speculating that if there is a positive correlation between X-ray luminosity and energy scale as was observed for other extreme HBLs such as Mrk 501 and 1ES 2344+514, the synchrotron peak might shift to even higher energies during flares.

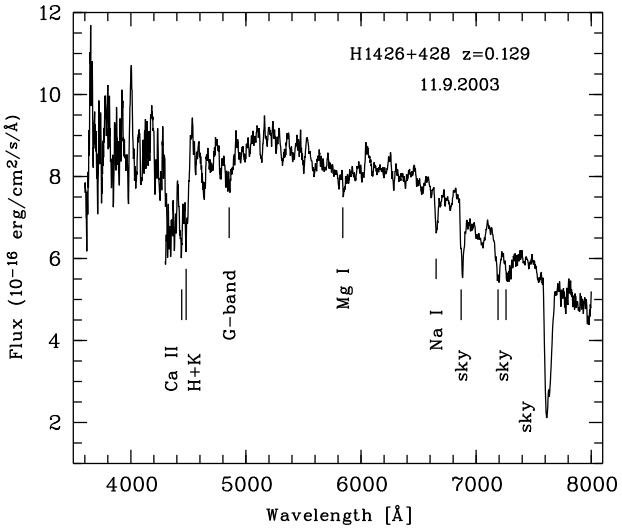

Figure 2.17: Optical spectrum of the host galaxy of H 1426+428, taken in 2003 with the CAFOS spectrograph of the 2.2 m reflector at Calar Alto. The prominent absorption lines were used to verify the blazar's redshift z = 0.129. From Götting (2006).

observed by SWIFT (March-June 2005; Tramacere *et al.*, 2007) and INTEGRAL (May 2006; Wolter *et al.*, 2007). The former group found that fitting log-parabolic fit functions to their data suggested synchrotron peak energies between 1 and 2 keV, depending on brightness state. Due to SWIFT's rather narrow energy range (0.6 – 9 keV), however, a completely flat spectrum (*i.e.* with spectral index $\alpha = 2$) to beyond 8 keV cannot be ruled out by these data. Furthermore, since contemporaneous data from SWIFT's UV detector was included in the fit, the ability of the used fit function to describe the SED over such a large energy range is questionable. INTEGRAL observations revealed a signal from H 1426+428 in the two X-ray detectors JEM-X and IBIS(ISGRI), which together cover the energy range from $\sim$3 keV to 10 MeV. The spectrum that was obtained is consistent with $E_{\rm peak} > 100$ keV, thus supporting the BeppoSAX results indicating that H 1426+428 is in fact an extreme high energy peaked BL Lac (Wolter *et al.*, 2007).

Blazars having the synchrotron peak at such high energies are very rare. Presently, H 1426+428 is one of a fairly small group of BL Lac objects referred to as "extreme synchrotron blazars"[40], which stand out through very high synchrotron peak energies. Even among these, Mrk 501 and 1ES 2344+514 are the only two BL Lacs that have also been observed in states with $E_{\rm peak} \gtrsim 100$ keV. An SSC model fit of H 1426+428's broad-

[40]Other blazars falling into this category are Mrk 501, 1ES 2344+514, 1ES 1959+650, PKS 2155-304, PKS 0548-322, 1ES 0033+595, 1ES 0120+340, 1ES 1218+304, H 2356-309, 1ES 1101-232, 1ES 0229+200 and 1ES 0347-121.

band spectrum (see Sec. 2.6.5) suggests values of $E_{\rm peak}$ that are higher than those of all other hitherto observed blazars (Wolter *et al.*, 2007).

Another interesting property of this HBL is the fact that the position of $E_{\rm peak}$ appeared to be independent of X-ray flux state, and was sometimes found to be in the 100 keV range even during periods of low activity, *i.e.* with the source in a quiescent state (Costamante *et al.*, 2001). Such behaviour has so far not been seen in other BL Lacs.

2.6.4 Correlations between X-rays and gamma rays

As we have seen in Sec. 2.3, all existing gamma-ray emission models for blazars include an interconnection between the synchrotron and high-energy peaks of the SED. The leptonic emission models, in particular SSC, call for tight correlations between synchrotron and Inverse-Compton fluxes, since the same accelerated electrons are responsible for emitting the synchrotron photons and up-scattering low-energy photons to gamma-ray energies. In a search for correlations between X-ray and gamma-ray fluxes of H 1426+428 on longer time scales, Aharonian *et al.* (2003) compared the average X-ray flux levels recorded by the *All-Sky Monitor* (ASM) detector on board RXTE (see Levine *et al.*, 1996) that were contemporaneous with the observation periods of the various gamma-ray experiments mentioned in the Sec. 2.6.2. The conclusion was that the rates were compatible with each other for the observing seasons of 1998, 1999 and 2000 (see Table 2.2 for a summary). In 2001, however, the X-ray flux appeared to have decreased by a factor of four, and during HEGRA's 2002 observing campaign it was still a factor of 1.7 lower than in 1999/2000. Thus the X-ray and gamma-ray (see Sec. 2.6.2) fluxes both seem to have decreased substantially between the 2000 and 2002 observing periods, suggesting a *positive* correlation between the X-ray and gamma-ray fluxes.

γ-ray observations	ASM avg. integr. X-ray flux [Hz]
CAT 1998/1999/2000	0.20 ± 0.40
Whipple 2001	0.26 ± 0.07
HEGRA 1999/2000	0.33 ± 0.08
HEGRA 2002	0.19 ± 0.04

Table 2.2: Average count rates of the ASM X-ray all-sky monitor detector on board RXTE. The numbers were obtained by averaging over all 90-second ASM measurements for all days on which gamma-ray observations were made. Included were only those observations from which energy spectra were deduced. From Horns *et al.* (2003).

Observations with RXTE-ASM are *unpointed*, in so far as this X-ray spectrometer with its large field of view ($6° \times 90°$) is intended to survey large sectors of the sky at the same time. Consequently, its sensitivity to point sources is low: averaged over 9 years, the typical count rate from the direction of H 1426+428 was (2.62 ± 0.07) mCrab (0.2 Hz, compared to 75 Hz for the Crab Nebula; Götting, 2006). This is below ASM's estimated sensitivity for daily averages (10 mCrab; Levine *et al.*, 1996), and causes even monthly averages to have large statistical errors. In 2000, 2001 and 2002, H 1426+428 was also observed with the pointed PCA spectrometer of RXTE, which has much higher sensitivity than ASM. Unfortunately these observations were not contemporaneous with any gamma-ray observations, yet it is still interesting to note that the measured X-ray flux was almost 50% *higher* in 2002 than it was in 2000, with a level of confidence of 7.8σ (Horns *et al.*, 2004). Naturally, due to the non-simultaneity with gamma-ray observations, an *anti*-correlation cannot be concluded, except if one allowed only a very gradual change of the emission spectrum over several months. For all PCA observations the spectrum between 2 and 20 keV remained hard, yet slightly harder when the flux was higher (spectral index $\simeq 1.8$ in high, and $\simeq 2.0$ in low states; Horns *et al.*, 2003).

A comparison of the Whipple data of 2000 and 2001 to contemporaneous ASM rates by Horan *et al.* (2002) also seemed to suggest an anti-correlation between the two when averaged over intervals of a month. Averaged over the whole observing period, the overall fluxes of the two instruments appeared to be *positively* correlated (a 60 % rate increase in ASM compared to an increase from $3.1\,\sigma$ to $5.5\,\sigma$ in the Whipple flux). The monthly ASM averages have large uncertainties and no statistical test was performed, thus the hint of a negative correlation may in this case be due to random fluctuations.

Finally, according to Smith *et al.* (2006), the average ASM X-ray flux during the CELESTE observations in the spring of 2004 was (0.2 ± 0.1) counts per second, suggesting a period of low X-ray activity. Under the assumption that the ASM count rate is a sufficiently reliable indicator of the source activity, one can assume that the blazar was in a quiescent state during the CELESTE observations, which could be an explanation for this experiment's non-detection of the source around 100 GeV.

2.6.5 Emission models for H 1426+428

Several authors have modelled the broad-band SED of H 1426+428 with single-zone SSC and SSC/EC gamma-ray emission models, using a variety of different EBL shapes to correct the VHE spectra for absorption (see *e.g.* Costamante *et al.*, 2003; Kato *et al.*, 2006; Wolter *et al.*, 2007). For completeness, the range of the fitted parameters are

as follows: radius of the emission region $R = (0.8-2.5)\cdot 10^{16}$ cm, magnetic field $B = 0.08-0.53$ G, bulk Lorentz factor $\Gamma = 17-20$, observing angle $\theta = 1.8°-2.0°$, Doppler factor $\delta = 25.1-27.3$, peak and maximum Lorentz factors $\gamma_{\rm peak} = (0.7-5.8)\cdot 10^6$ and $\gamma_{\rm max} = (3-40)\cdot 10^6$, and injection spectral index $s = 1.9-2.9$ (parameters are explained in Sec. 2.3.1). In the leptonic models, on which we shall focus in this section, the energies of the synchrotron photons are directly related to the energies of the accelerated electrons in the jet by which they are emitted. A high synchrotron peak energy is thus indicative of high electron peak Lorentz factor (see Eq. 2.33). In Fig. 2.18, the fitted peak Lorentz factors for most of the gamma-ray detected blazars are plotted over energy density, showing the typical anti-correlation that defines the blazar sequence (Ghisellini *et al.*, 2002). As we can see, the result obtained for H 1426+428 in this particular model (Wolter *et al.*, 2007) yields a peak Lorentz factor $\gamma_{\rm peak} \simeq 3\cdot 10^6$, and confirms the physical interpretation of the blazar sequence at the very high-energy end.

The distribution of Lorentz factors of the accelerated electrons, in turn, translate directly into the energies of the up-scattered target photons, independent of whether the latter originate from synchrotron emission (SSC) or from outside the jet (EC). According to this idea, and from what was said about the synchrotron peak in Sec. 2.6.3, it then follows that the peak of the high-energy (IC) hump of this blazar should – at least during high states – be located at extremely high energies, *i.e.* in the TeV range. Indeed, the results of Costamante *et al.* (2003) suggest that the IC-peak is located above $\sim$10 to 12 TeV (see also Kneiske *et al.*, 2004). Due to its relatively large redshift however, VHE gamma rays from H 1426+428 are strongly attenuated through collisions with the extragalactic infrared background radiation, and the observed spectrum is therefore expected to be heavily modified (see Sec. 2.5.4). In particular, it is not expected to be a power law, but cut off towards higher energies with a shape determined by the EBL spectrum at optical to mid-IR wavelengths. Due to the large optical depth at TeV energies in particular, a spectral peak at these energies would also remain hidden from the observer.

From what was said in Sec. 2.5.4, insufficient knowledge of the EBL density currently prevents any detailed modelling of blazars with higher redshifts. For instance, different assumptions about the normalisation and shape of the EBL spectrum lead to very different shapes and locations of the intrinsic high-energy peaks of large-redshift blazars. In the case of H 1426+428 this dependence is particularly pronounced, since this source does not seem to be well fitted by current SSC/EC models, and thus the high-energy hump cannot readily be inferred from the synchrotron component of the SED like for some other blazars. This becomes evident when the observed VHE spectrum is cor-

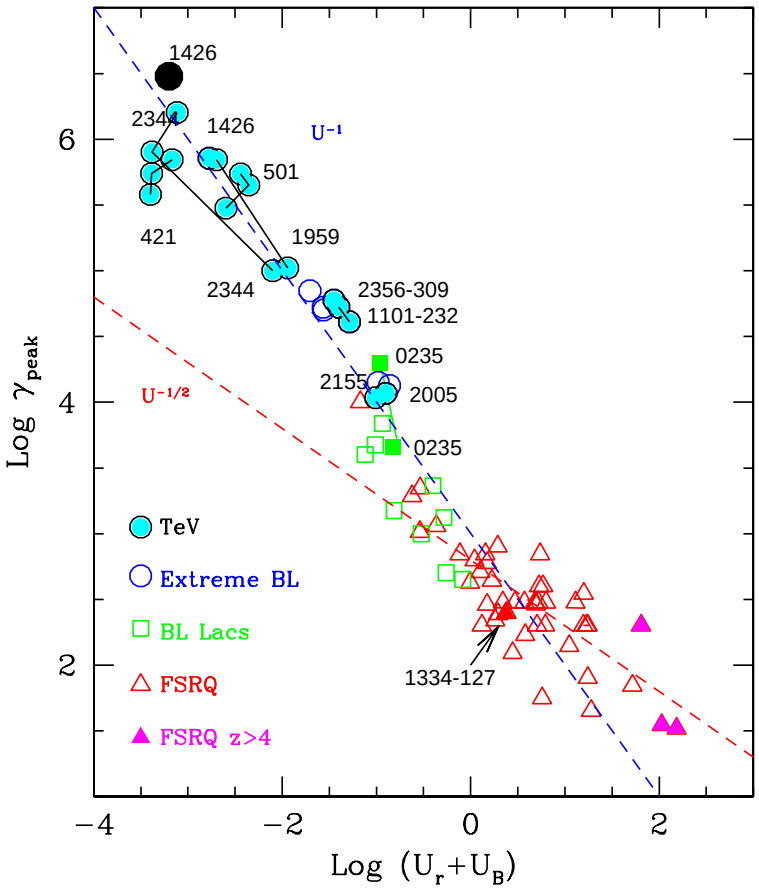

Figure 2.18: Lorentz factors $\gamma_{\rm peak}$ of the electrons emitting at the peaks of the SED (according to the SSC model) as a function of co-moving energy density (magnetic plus radiative), for a comprehensive blazar ensemble. The fit parameters of H 1426+428, denoted by the filled black circle, populate the extreme high-energy end of the sequence. Note that slightly further down the sequence is a second point for H 1426+428, which corresponds to a different fit using only older X-ray data. Taken from Wolter *et al.* (2007), adapted from Ghisellini *et al.* (2002).

rected for absorption in the EBL and fitted with a single-zone SSC (or SSC/EC) model. Intrinsic spectra obtained in various works (see *e.g.* Costamante *et al.*, 2003; Mapelli, Salvaterra & Ferrara, 2004), assuming EBL models with medium to high overall flux level (see Sec. 2.5), suggest an unusually high ratio of IC to synchrotron luminosity ($L_{\rm IC}/L_{\rm sync} \approx 10$) combined with an IC-peak at $\rm E_{IC} \gtrsim 10$ TeV. This however would violate the blazar sequence scenario presented in Sec. 2.4.3. More specifically, the SSC model alone does not seem to be able to account for the high luminosity of the IC component, requiring an additional external seed radiation field to be present (see Sec. 2.3.1). A significant EC contribution, in turn, is quite untypical for a HBL, let alone such an

extreme one, as external radiation also tends to hinder the electron acceleration (hence the expected anti-correlation of peak energy and IC-luminosity, which constitutes the blazar sequence). If, on the other hand, EBL models with low optical/near-IR fluxes are used, which are in agreement with recent upper limits obtained from the spectra of several high-redshift HBLs (Sec. 2.5.4; Dwek & Krennrich, 2005; Aharonian *et al.*, 2006b; Costamante, 2007; Mazin & Raue, 2007), the derived intrinsic spectrum is sufficiently softened so that the IC-peak shifts to much lower energies. At the same time dominance of the high-energy component is reduced (see *e.g.* Kato *et al.*, 2006; Wolter *et al.*, 2007).[41]

The broad-band SED of H 1426+428 is shown in Fig. 2.19, including measured fluxes from radio to TeV energies. It also shows corrected VHE fluxes for two different EBL shapes: a higher (deemed more realistic) and a lower EBL flux model, both of Kneiske *et al.* (2004). An SSC model fit to the VHE fluxes which were "de-absorbed" with the lower flux model is also shown. In the X-ray, the unusually broad synchrotron peak which is characteristic for this blazar can clearly be seen. Even though the EBL parameterization used is probably too low, as it does not account for dust-obscured (so-called infrared) galaxies (Kneiske, Mannheim & Hartmann, 2002), the $\nu\,\mathrm{F}_\nu$ value obtained at the IC peak is almost a factor of 10 higher than that at the synchrotron peak. Note that single-zone SSC models are restricted to modelling compact emission regions in the jet, and hence do not account for the radio emission properly. Therefore the radio fluxes are not included in the model fit.

The most obvious conclusion from the study of the SED of H 1426+428 is that the EBL flux appears to be low, *i.e.* not much higher than the lower limits from galaxy counts, and in accordance with the upper limits from the blazar studies mentioned in Sec. 2.5.4. As already explained there however, the validity of the postulated physical upper limit on the hardness of intrinsic blazar spectra, upon which these upper limits are based, is questionable. Apart from the arguments mentioned in Sec. 2.5.4, there could be other physical processes present which give rise to harder spectra than permitted by the basic SSC/EC model. As part of the latter possibility, an explanation for very high fluxes at TeV energies could be a significant contribution from accelerated hadrons (Sec. 2.3.2). In that case, however, it is difficult to justify why the jets of H 1426+428 – as well as most of the other farther HBLs, whose intrinsic spectra in this case would also

[41] Using a low-flux EBL parameterization similar to Dwek & Krennrich (2005) and assuming a synchrotron peak energy of around 40 keV – corresponding to a low state of the source – Kato *et al.* (2006) obtained an IC-peak energy of only 85 GeV when fitting the SED with a pure SSC model. For the observed "high" states, when $E_{\mathrm{peak}} \gtrsim 100$ keV, the IC hump is of course expected to peak at energies higher than that.

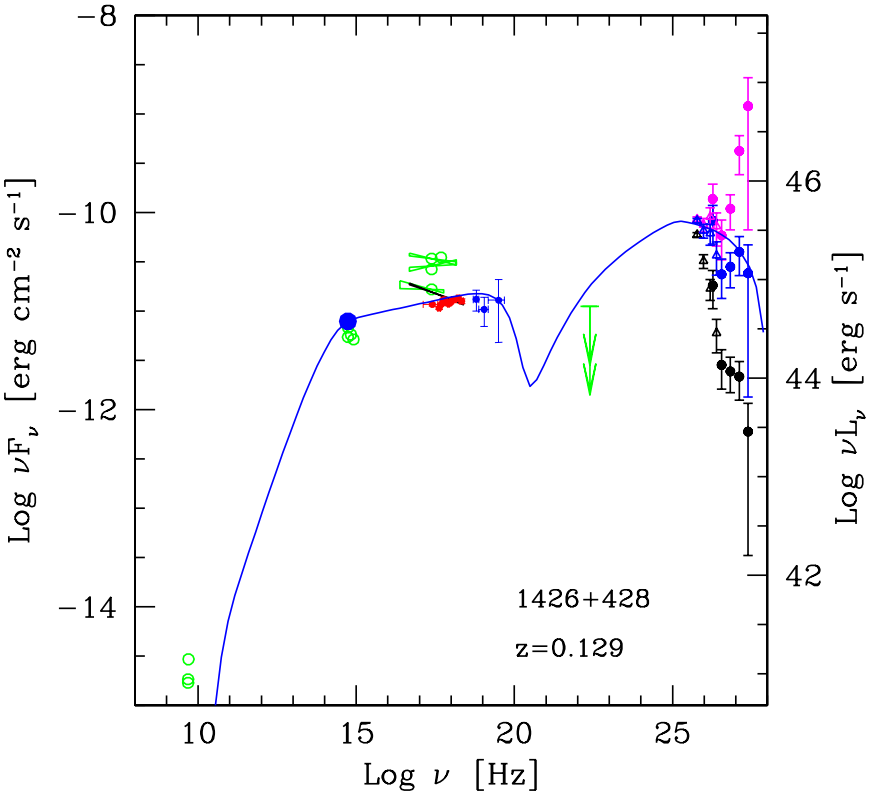

Figure 2.19: Broad-band SED of H 1426+428, including an example of a pure SSC model fit (solid line, from Wolter *et al.*, 2007). The shown measurements (for the most part not contemporaneous) are as follows. Radio and older optical observations (*empty green circles*), optical observations from May 2006 (*large blue circle*). X-ray: BeppoSAX/MECS (Feb. 1999; *small red circles*), ASCA and ROSAT (*empty green circles with butterflies*), INTEGRAL/ISGRI (May 2006; *small blue circles*). Gamma rays: EGRET UL (*green arrow*), CAT (1998-2000; *black empty triangles*), HEGRA (1999-2000; *black circles*). The VHE measurements have been absorption-corrected by means of two different EBL parameterizations from Kneiske, Mannheim & Hartmann (2002), which are both shown: "best fit" model (*magenta circles*) and "low-IR" model (*blue circles*). It should be noted that in particular the high-energy peak of this blazar is not well described by a pure SSC model. As a representative example, the model fit shown fails to agree with the first (more realistic) EBL-model corrected VHE fluxes, while possibly overestimating the flux at EGRET energies. Yet the ratio of IC to synchrotron luminosity is unusually high for a HBL. Taken from Wolter *et al.* (2007).

be very hard – should contain a larger hadronic component than closer VHE gamma-ray emitting HBLs such as Mrk 421 and Mrk 501.

Considering the fact that the more popular EBL parameterizations appear to favour steadily rising intrinsic spectra beyond 1 TeV for all of the more distant HBLs, one could

imagine that perhaps the *de facto* unknown EBL-flux in the "valley" between near- and far-IR is consistently being overestimated.

Last but not least it should be kept in mind that in as far as model detail is concerned, basically all of the SSC models used for blazar studies up to now are single-zone models (see Sec. 2.3.1). It is thus conceivable that some of the underlying assumptions are oversimplifications of reality, and that by increasing detail and adding inhomogeneities (*e.g.* regarding the distribution of the magnetic field and the emission regions), some of the apparent discrepancies between leptonic models and absorption-corrected data can be resolved. New observations of H 1426+428 with the current generation of ACTs (MAGIC and VERITAS), ideally as part of a multi-wavelength campaign, should be able to constrain possible emission models, and also help answering the question as to how the broad-band SED changes for different flux states, which appears to be unusual in this source (see Sec. 2.6.4).

2.6.6 Implications for the Extragalactic Background Light

Due to its redshift being significantly larger than that of all previously detected extragalactic VHE gamma-ray emitters, the TeV spectrum of H 1426+428 measured by HEGRA provided the first observation of an imprint of the EBL in the SED of a blazar's gamma-ray spectrum (Aharonian *et al.*, 2003). Various authors have used the VHE spectrum of H 1426+428 to constrain the EBL fluxes in the near- to mid-infrared, assuming that the intrinsic spectrum cannot be harder than $\alpha = 1.5$ (see Sec. 2.5.4). As of January 2008, the source is one of only five blazars for which fluxes up to energies of above 10 TeV have been measured, and one of only two of these which are located at redshifts larger than 0.1 (the other being the recently observed HBL 1ES 0229+200; see Aharonian *et al.*, 2007b).[42] Only gamma-ray fluxes at energies above a few TeV are affected by EBL photons of the poorly explored mid-IR band. Provided our knowledge about intrinsic blazar spectra improves, the HEGRA data for the relatively distant H 1426+428 could therefore be used for putting constraints on that part of the EBL (see *e.g.* Mazin & Raue, 2007).

[42] Ironically, the fact that the new-generation ACTs have much lower energy thresholds than the pioneer imaging telescopes has led to a change in observing strategies, so that rarely enough time is spent observing one particular source to collect a sufficiently large sample of multi-TeV events, as it was the case for *e.g.* HEGRA and Whipple. The reason why the measured spectrum of 1ES 0229+200 still extends up to 15 TeV is that at the latitude of HESS, this source culminates at a very large zenith angle (43.6°), resulting in a high energy threshold of 580 GeV and at the same time very high effective areas in the TeV range.

Although the next section is not specifically devoted to H 1426+428, we shall return to the discussion of this source with some new insights in its conclusion (Sec. 2.7.3).

2.7 Secondary gamma-ray emission

As mentioned in Sections 2.5 and 2.6.6, VHE gamma-ray spectra from distant sources can be used to probe the extragalactic background radiation (EBL). The vast majority of studies dealing with the effects of EBL-absorption on gamma-ray spectra from extragalactic sources merely subtract the fraction of primary gamma rays that are thought to collide with EBL photons from the intrinsic spectra. This way of reasoning, however, is strictly correct only if one considers solely those gamma rays coming from the *exact* direction of the source. That may seem adequate under the assumption that secondary particles resulting from these collisions are either not beamed towards the observer, and/or do not contain or give rise to gamma rays. As will be argued in the following, neglecting the secondary processes is generally justifiable for the part of the SED above $\sim 1\,\mathrm{TeV}$, but, depending on the (as yet unknown) strength of the intergalactic magnetic field, not for the part in the GeV energy range. We will now direct our attention at the products of the absorption process, following a line of argumentation presented by Dai, Zhang *et al.* (2002), Fan *et al.* (2004) and Neronov & Semikoz (2006).

2.7.1 Deflection angle and secondary gamma-ray energy

Primary gamma rays with energies in the TeV-range pair-produce through collision with infrared photons of the EBL after a mean free path of typically a few tens of Mpc (see Eq. 2.42). An important prerequisite for the following argument is the fact that in most scenarios, this distance is large enough to allow the gamma rays to escape from the presumably high magnetic fields surrounding the source ($\sim 10^{-9}\,\mathrm{G}$ for blazars), causing the actual collision to occur in the intergalactic voids, *i.e.* regions filled only with the much lower primordial *intergalactic magnetic field* (IGMF).[43] If E_γ is the energy of the primary gamma ray and m_e the electron rest mass, the Lorentz factors γ_e of the resulting electron-positron pair will then be

$$\gamma_e \simeq \frac{E_\gamma}{2m_e c^2} \simeq 10^6 \frac{E_\gamma}{1\,\mathrm{TeV}} \tag{2.50}$$

[43] The IGMF, often also called EMF (*Extragalactic Magnetic Field*), is thought to follow the observed filamentary large scale structure (LSS) of the universe. Hence it is expected to be lowest in the voids, and somewhat higher in the filaments that connect the galaxy clusters.

Keeping in mind that all observed energies are redshifted, we can express E_γ in above equation in terms of the *observed* gamma-ray energy $E'_\gamma = (1+z)^{-1}E_\gamma$, where z is the redshift of the source, yielding $\gamma_e \simeq (1+z)E'_\gamma/(2m_ec^2)$.

Extragalactic space is permeated by photons with wavelengths in the millimetre-range[44] belonging to the Cosmic Microwave Background radiation. Thus, after a typical distance of

$$d_{\rm IC} \;=\; c\,t_{\rm IC} \;\simeq\; c\,\frac{3\,m_e\,c}{4\,\gamma_e\,\sigma_{\rm T}\,u_{\rm CMB}} \;\simeq\; 2.3\cdot 10^{24}\,\frac{10^6}{\gamma_e}\,(1+z)^{-4}\ {\rm cm} \qquad (2.51)$$

the electrons and positrons collide with CMB photons, inverse-Compton scattering them to GeV energies. Here $t_{\rm IC}$ is the IC cooling time scale of an electron/positron with Lorentz factor γ_e as measured in the local rest frame[45], $\sigma_{\rm T}$ the Thomson cross-section as defined in Eq. 2.26, and $u_{\rm CMB} = aT^4$ ($= 4.72\cdot 10^{-3}$ eV cm^{-3} K^{-4} $\cdot\,(1+z)^4\,\cdot$ (2.73 K)^4) the CMB energy density at redshift z.[46] From Eqns. 2.51 and 2.50 it is evident that the mean free path of the e^+/e^- pair is much shorter than the distance of the (extragalactic) source to the observer, thus the energy loss due to IC can be considered to be local.

It can be shown that the angle $\theta_{\rm IC}$ between the ultra-relativistic e^+/e^- and the up-scattered photon in the local rest frame is in the order of $\sim 1/\gamma_e^2$, and thus negligible. While travelling the distance $d_{\rm IC}$, however, electron and positron are subject to deflection by the intergalactic magnetic field $B_{\rm IG}$.[47] The latter is thought to be more or less randomly oriented, so that to first order there is no preferred direction (Kneiske & Mannheim, 2005) and the magnitude of the deflection can be estimated by using the relativistic Larmor radius $R_{\rm L}$ of the electron/positron, which according to Eq. 2.15 is, in Gaussian units,

$$R_{\rm L} \;=\; \frac{\beta\,\gamma_e\,m_e\,c^2}{e\,B_{\rm IG}} \qquad (2.52)$$

with e being the electron charge. Since the e^+/e^- are ultra-relativistic, we can set $\beta \simeq 1$. Using Eqns. 2.51 and 2.52, it can be verified that $R{\rm L} \gg d_{\rm IC}$, and thus the maximum

[44] In the present epoch, the peak of the CMB's blackbody spectrum is located at a frequency of about 160 GHz, corresponding to a wavelength of $\lambda_{\rm CMB} = 1.9$ mm.

[45] Here the term "local rest frame" refers to the frame in which the dipole moment of the CMBR vanishes, *i.e.* to first order the rest frame of the gamma-ray source.

[46] Strictly speaking the redshift correction for the CMB temperature should be somewhat smaller than this, since the CMB scattering does not occur right at the source but a fair distance away towards the observer.

[47] Due to the low field strength, the energy loss due to synchrotron radiation is quasi non-existent.

deflection angle $\theta_{\rm B}$ can readily be estimated via

$$\theta_{\rm B} \simeq \frac{d_{\rm IC}}{R_{\rm L}} = \frac{3\,e\,B_{\rm IG}}{4\,\gamma_e^2\,\sigma_{\rm T}\,u_{\rm CMB}} \tag{2.53}$$

Substituting Eqns. 2.51 and 2.50 in above, we arrive at

$$\theta_{\rm B} \simeq 1.35 \cdot 10^{15}\,(1+z)^{-4}\,\left(\frac{10^6}{\gamma_e}\right)^2\,\left(\frac{B_{\rm IG}}{1\,{\rm G}}\right) \tag{2.54}$$

$$= 1.41 \cdot 10^{15}\,(1+z)^{-6}\,\left(\frac{\rm TeV}{E'_\gamma}\right)^2\,\left(\frac{B_{\rm IG}}{1\,{\rm G}}\right) \tag{2.55}$$

This simple expression relates the maximum deflection angle $\theta_{\rm B}$ directly to the redshift-corrected energy of the primary photon (*i.e.* the energy it would have at the observer location if it had not been absorbed) in TeV and the intergalactic magnetic field strength in Gauss.

The second quantity of interest is the energy $E_{\rm sec}$ of the up-scattered CMB-photon (or *secondary* photon). Since the collision of TeV electrons with microwave photons is governed by Thomson kinematics, $E_{\rm sec}$ is proportional to γ_e^2 (see Eq. 2.29) and thus to the square of the primary gamma ray's energy. Averaged over the angular distribution, the expression for the energy is (Neronov & Semikoz, 2006)

$$E_{\rm sec} \simeq \frac{4}{3}\,\gamma_e^2\,E_{\rm CMB} = (1+z)^2\,\frac{(E'_\gamma)^2\,E_{\rm CMB}}{3\,m_e^2\,c^4} \tag{2.56}$$

where we again applied the redshift-corrected version of Eq. 2.50. Note that technically the energy of the CMB photon would have to be redshift-corrected as well through multiplication with $(1+\tilde{z})$, where $\tilde{z}$ $(< z)$ is the redshift at the time of the pair-production process. However, this factor is cancelled out by the inverse of the same factor due to the cosmological redshift to which the secondary photon is subjected on the way from the pair-production site to the observer. If we take $E_{\rm CMB}$ to be the mean energy of the CMB blackbody spectrum at present epoch, *i.e.* $E_{\rm CMB} \simeq 2.7\,k\,T = 2.7k \cdot 2.73\,{\rm K}$ $\simeq 6.6 \cdot 10^{-4}\,{\rm eV}$, the numerical dependence of the secondary gamma ray's energy on the primary energy according to Eq. 2.56 is

$$E_{\rm sec} \simeq 0.84\,\left(\frac{E'_\gamma}{\rm TeV}\right)^2\,(1+z)^2\ {\rm GeV} \tag{2.57}$$

Thus a source emitting gamma rays above 10 TeV of energy will, via interaction with the EBL, produce secondary gamma rays above 100 GeV.

The ratio of the post-scattering energy of the ultra-relativistic electron/positron ($\tilde{E}_\mathrm{e}$) to its energy before the IC-scattering process (E_e) is bounded by

$$\frac{\tilde{E}_\mathrm{e}}{E_\mathrm{e}} \gtrsim \sqrt{1 - 8\,\gamma_e \frac{E_\mathrm{CMB}}{m_e c^2}} \tag{2.58}$$

corresponding to the case of maximum momentum transfer (head-on collision). In the Thomson limit, for which the above equation was derived, $\gamma_e^2 \ll m_e c^2/E_\mathrm{CMB}$; therefore the energy loss through IC is small enough to enable one $\mathrm{e}^+/\mathrm{e}^-$ to scatter up to 1000 CMB photons. Furthermore, the recoil of the $\mathrm{e}^\pm$ is sufficiently small so that the deflection angle due to multiple scattering after the N^th IC-collision is only $\theta_\mathrm{N} \approx \sqrt{N} E_\mathrm{CMB}/m_e c^2$. For $N \approx 1000$, $\theta_\mathrm{N} \approx 10^{-7}$, and thus the whole cascade is strongly beamed in the direction of the projectile. The resulting two "jets" of secondary gamma rays – one due to each member of the $\mathrm{e}^\pm$-pair – are, under the influence of the IGMF, curved away slightly from the direction of the original gamma ray in opposite directions. Despite the fact that it does not result from an electromagnetic cascade in the strict sense, this phenomenon is also referred to as *cascade emission*. With increasing number of scatters, of course, the net deflection of the emitted cascade photons increases; the number of secondary photons beamed in a direction sufficiently close to that of the first up-scattered photon depends on the strength and direction of the local magnetic field and on the primary energy.

Last but not least, the *time scale* of the secondary gamma rays plays a role when the source is variable, such as in the case of blazars. While the time delay due to the brief electron/positron phase of the secondary radiation is negligible, the path length differences resulting from the magnetic deflection are, due to the large distances involved, substantial even for tiny angles. The arrival of the secondary radiation is delayed with respect to the primary gamma rays by an estimated

$$\Delta t_\mathrm{obs} \approx \frac{1}{2}(1+z)\, t_\mathrm{IC}\, \theta_\mathrm{B}^2 \tag{2.59}$$

(see *e.g.* Plaga, 1995; Fan *et al.*, 2004), where t_IC is the Inverse-Compton time scale given in Eq. 2.51 and θ_B the magnetic deflection angle. For a primary gamma-ray energy of 10 TeV, the expected time delay is in the order of seconds or less if the magnetic field strength is below 10^{-19} G, while it is tens of years if the field strength is $\sim 10^{-15}$ G. In principle, this retardation of the secondary radiation can be used to estimate the strength of the IGMF, provided the temporal origin of the primary radiation is well defined and the source redshift is known. For instance, applying Eqns. 2.59 and 2.55 to the delayed ($\Delta t_\mathrm{obs} \sim 5000$ s) MeV-emission of GRB 940217 ($z \approx 0.68$) suggests a weak

IGMF of $B_{\rm IG} \sim 10^{-20}$ G (Fan *et al.*, 2004). Since the deflection angles and corresponding time delays are more or less random within a cone of half-angle $\theta_{\rm B}$, variability of blazar primary emission will only show up in the secondary emission for very small values of the IGMF ($\lesssim 10^{-17}$ G), dampened and with a time delay. For higher field strengths, the cascade emission is expected to be steady and independent of source activity.

2.7.2 Detection of secondary emission

In the preceding it has been explained that the absorption of VHE gamma rays in the EBL produces secondary gamma rays which are emitted at angles $\theta \lesssim \theta_{\rm B}$ with respect to the primary direction. Therefore, if the beam of primary gamma rays were infinitely fine and emitted exactly in the direction of the observer, no secondary radiation would arrive there. We know however – beginning with the mere fact that gamma-ray emission from numerous blazars is detected which must be relativistically beamed towards Earth – that the relativistic jets of blazars have opening angles δ of several degrees.[48] It is therefore possible for secondary gamma rays, which are produced by the EBL-absorption of primary TeV-photons emitted at some small angle ϕ with respect to the line of sight, to be "refocussed" onto the observer. A sketch of the underlying principle is shown in Fig. 2.20.

From simple geometry, the relationship between the maximum deflection angle $\theta_{\rm B}$ and the maximum observed arrival direction ρ with respect to the source direction can be expressed in terms of the optical depth τ:

$$\sin\rho \;=\; \frac{1}{\tau(E_\gamma, z)}\,\sin\theta_{\rm B}(B_{\rm IG}, E_\gamma) \tag{2.60}$$

Note that due to the random orientation of the IGMF that is traversed, $\theta_{\rm B}$ is just an upper bound for the actual deflection angle. Together with the fact that $d_{\gamma\gamma}$ is just the mean value of the actual pair-production distances of individual gamma rays, this means that the observed arrival angles are distributed more or less evenly within a solid angle of half-angle ρ, forming a halo. Another contribution to the smearing effect of the secondary radiation arrival direction is the fact that one electron-positron pair up-scatters many CMB photons to gamma-ray energies before it gets strongly deflected (Sec. 2.7.1), and the cascade due to subsequently up-scattered CMB photons has a finite opening angle.

[48] Common models estimate δ to be around 6° (see *e.g.* Urry & Padovani, 1995).

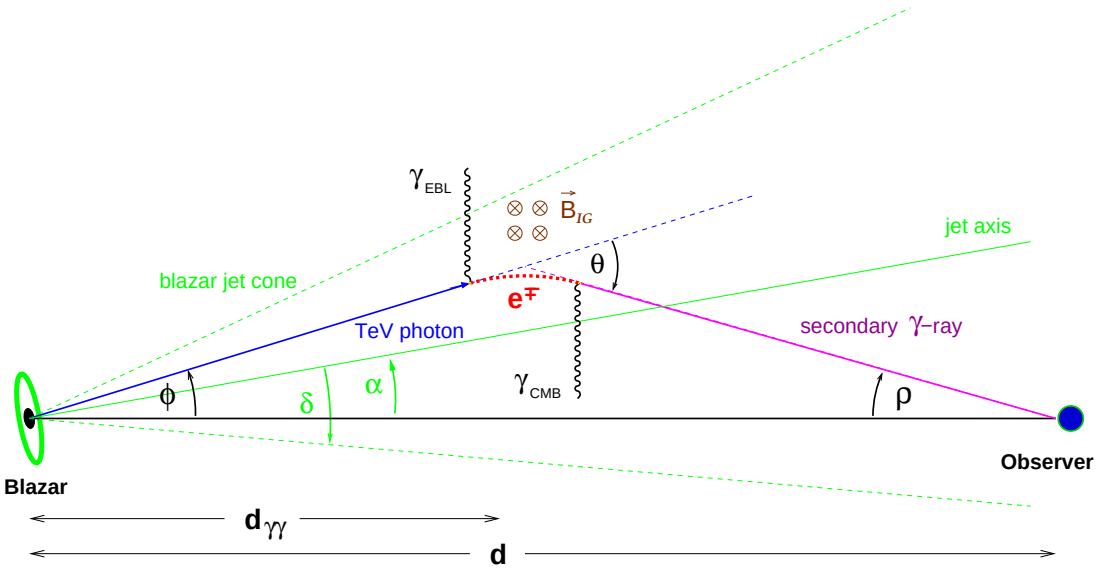

Figure 2.20: Sketch of the principle of secondary (or *cascade*) gamma-ray detection. A blazar at a distance d, whose jet axis is offset from the line of sight by some small angle α (often named θ in the literature), emits TeV gamma rays within a cone of opening angle δ. After a mean distance $d_{\gamma\gamma}$ the TeV photons pair-produce in the EBL. The resulting electrons/positrons are deflected in the IGMF by an average angle θ before they up-scatter CMB photons to GeV energies via Inverse Compton effect. For any given primary energy and magnetic field strength $B_{\rm IG}$, there exists one emission angle ϕ for which the secondary photons are directed towards the observer, arriving at a small angle ρ with respect to the line of sight to the blazar (provided that $\delta > \phi/2$). All angles are exaggerated for easier viewing.

The downstream part of the electromagnetic cascade is eventually diffused by the magnetic field, and adds to the cosmic gamma-ray background in the form of a faint halo. This gamma-ray background was detected by EGRET between 100 MeV and ∼50 GeV, and now appears to be well explained by means of a superposition of secondary emission from unresolved blazars (Kneiske & Mannheim, 2005, 2008).

The observability of secondary radiation depends of course strongly on the strength of the magnetic field at the location where the pair production process takes place. In the extragalactic voids, the magnetic field (IGMF) remains de facto unknown to date; values anywhere between 10^{-29} and 10^{-9} G are considered possible (for details, see *e.g.* Fan *et al.*, 2004; Neronov & Semikoz, 2006, and references therein).[49] In theory, the

[49]The magnetic field within galaxies, estimated through Faraday rotation and radio synchrotron emission measurements, is thought to be in the order of 10^{-9} G. Knowing the strength of the *extragalactic* magnetic field is, among others, important for understanding the propagation of UHE cosmic rays: if the mean IGMF is weaker than about 10^{-9} G, cosmic rays of energy around or above 10^{20} eV should point back at their sources.

diffusing effect of the IGMF on the cascade emission of TeV-emitting blazars provides an independent method for measuring the strength of the latter (Neronov & Semikoz, 2006). If the IGMF is larger than a certain critical value $B_{\rm IG}^{max}$, which depends on the primary gamma-ray energy and the opening angle of the relativistic jet, even the secondary gamma rays from the very first interaction with the CMB will not reach the observer who, by definition, views the blazar from within a narrow cone around the centre of the relativistic jet. For primary energies in the TeV range, $B_{\rm IG}^{max} \sim 10^{-12}$ G. If the field strength is less than this value, a diffuse halo due to cascade emission, whose diameter is a measure of the field strength, should be detectable around the source. Finally, for field strengths below $B_{\rm IG}^{min} \sim 10^{-15}$–$10^{-18}$ G (the actual value depending on the primary energy and the angular acceptance of the detector), the cascade emission should not be discernible from the point source flux of unabsorbed gamma rays, and be superimposed on the latter. This is the case which interests us here, as if the secondary gamma ray flux is significant, we can expect the SEDs of more distant blazars to be modified through the appearance of an additional component in the multi-MeV to GeV range.

Using Eqns. 2.55, 2.57 and 2.60, the value of $B_{\rm IG}^{min}$ can readily be estimated. For the STACEE gamma-ray detector (see Chapter 4) with its large angular acceptance of 0.8° half-angle and an energy threshold of around 160 GeV (see Sections 5.5.3 and 7.3), and using a low-flux EBL model such as that of Primack *et al.* (2005), the maximum IGMF strength that would cause the secondary emission to add to the point-source flux is $B_{\rm IG}^{min} \approx 10^{-15}$ G.

2.7.3 Modification of observed gamma-ray spectra

The expected contribution of secondary emission to the gamma-ray spectrum of Mrk 501 was calculated by Dai, Zhang *et al.* (2002), with the result that when the blazar is in a flaring state, the secondary component should be detectable as a bump on top of the low-energy slope of the IC-peak predicted by the SSC model, as long as the mean strength of the IGMF is less than $\sim 10^{-17}$ G. Due to the low redshift of this blazar however, the enhancement superimposed on the IC component would merely lead to a broadening of the peak towards lower energies, and in such a way be unequivocally identifiable only between a few hundred MeV and ∼100 GeV. Unfortunately this energy range is part of the still existing gap between satellite-based and ground-based gamma-

ray detectors.[50] With its larger redshift and therefore higher gamma-ray absorption, a TeV gamma-ray emitter such as H 1426+428 is hence a more promising candidate for identifying the predicted cascade emission. Figure 2.21 shows the expected contribution of secondary gamma rays in the MeV-GeV energy range to the spectrum of H 1426+428 as calculated by Fan *et al.* (2004), using three different basic EBL parameterizations (leading to three different intrinsic VHE spectra), and for three different strengths of the IGMF. Apparent is the increase of the cascade flux with increasing hardness of the intrinsic spectrum, equivalent to a higher level of the EBL flux in the mid-infrared. In this context it should be noted that only the EBL model used in the leftmost panel (*a*) is marginally compatible with recent estimates (see Sec. 2.5), while both (*b*) and (*c*) are considered unrealistic. Thus the question whether secondary emission plays a significant role in the observed gamma-ray spectrum of H 1426+428 depends largely on the actual shape of the EBL spectrum.

Note that for predictions of the cascade contribution, the underlying assumption is that the IGMF is weak enough so that all secondary gamma rays are within the detector acceptance, which – for the energies and optical depths involved – is more than justified, as suggested by Eq. 2.60.

In Fig. 2.22, the expected gamma-ray spectra at the observer (absorbed spectrum, and absorbed spectrum with cascade contributions for several different magnetic field strengths), are plotted for a source at a slightly higher redshift (1ES 1101-232, z = 0.186). In contrast to the former example, this calculation is based on a *low*-flux EBL model, consistent with the upper limits of Aharonian *et al.* (2006b). Intrinsic spectra of power-law form with two different spectral indices, also shown, are assumed. The very different normalisations and shapes of the spectra with and without cascade contribution are evident. Noticeable is the overall steepening (softening) of the observed VHE spectrum caused by the flat bump from secondary emission accumulating at GeV energies – a fact that, if left unaccounted for, would lead one to infer overly soft intrinsic spectra (or inversely, EBL fluxes that are too high; see also d'Avezac *et al.*, 2007). Also eye-catching, especially in the plot for the harder intrinsic spectrum, is the fact that cascade emission is expected to provide a contribution to the observed gamma-ray flux at *any* energy in the HE-VHE band, *i.e.* even in the TeV range, provided the intrinsic spectrum extends to high enough energies (as required to provide the seeds for the observed secondary gamma rays). It should be remarked that the intrinsic spectrum assumed in the calculations for

[50] According to Dai, Zhang *et al.* (2002), there could be a measurable enhancement of the gamma-ray flux from Mrk 501 due to cascade emission even between 100 and $\sim$200 GeV, but only in case of a high EBL-flux *and* $B_{\rm IG} \lesssim 10^{-17}$ G.

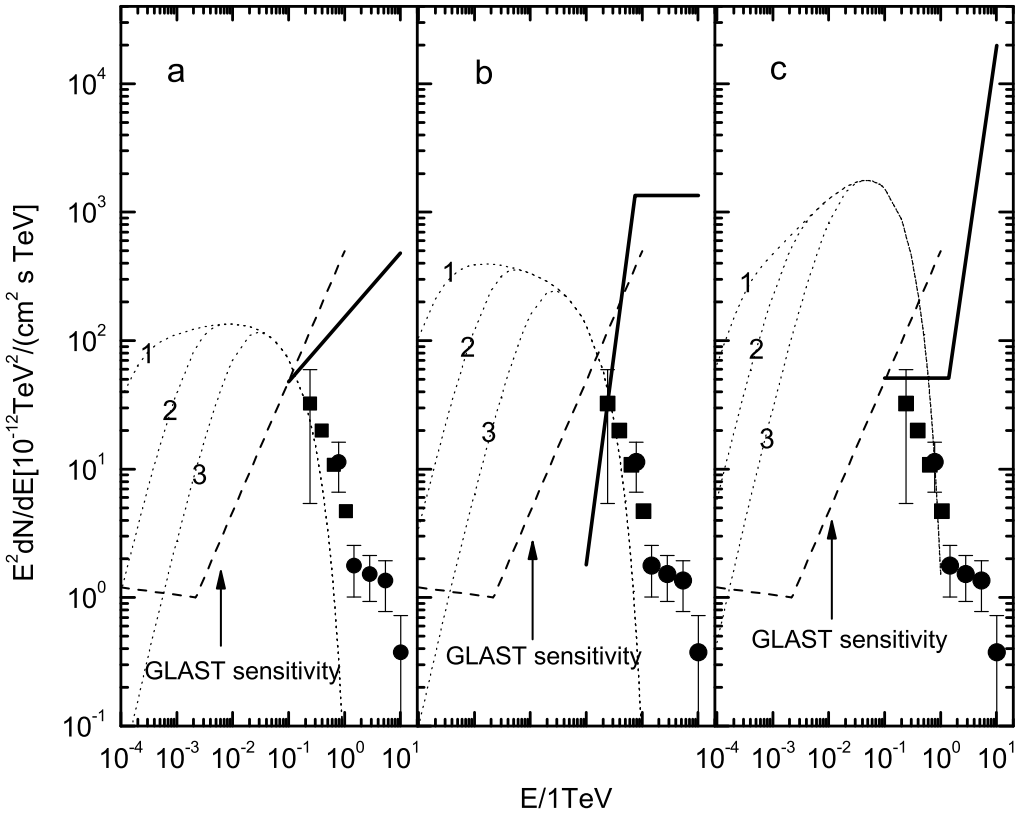

Figure 2.21: Predicted contribution of secondary (cascade) emission to the gamma-ray spectrum of H 1426+428, for three different assumed intrinsic spectral shapes (indicated by the solid line, which is based on the EBL correction models of (*a*) Primack, Somerville *et al.* (2001), (*b*) Aharonian *et al.* (2003), and (*c*) Malkan & Stecker (2001)), and three different strengths of the intergalactic magnetic field (dotted curves labelled 1 - 3): $B_{\rm IG} = 10^{-20}\,{\rm G}$ (1), $10^{-19}\,{\rm G}$ (2), and $10^{-18}\,{\rm G}$ (3). The expected sensitivity of GLAST (see Sec. 1.5.2) for an exposure time of 2 days is also plotted. Taken from Fan *et al.* (2004).

both plots of Fig. 2.22 extends to very high energies (1000 TeV) and as such is certainly unrealistic. For more probable cut-off values of between 10 and 30 TeV, the excess above $\sim 1\,{\rm TeV}$ is suppressed and the cascade contribution remains significant only in the GeV range (d'Avezac, 2006; d'Avezac *et al.*, 2007).

In summary, cascade emission can be expected to contribute measurably to the point-source flux of extragalactic sources if (in order of importance):

1. Pair production occurs in intergalactic voids, and the strength of the IGMF present there is very low ($B_{\rm IG} \lesssim 10^{-16}$ G).

2. The source's intrinsic emission spectrum is harder than $\alpha \lesssim 2.0$ and extends to sufficiently high energies.

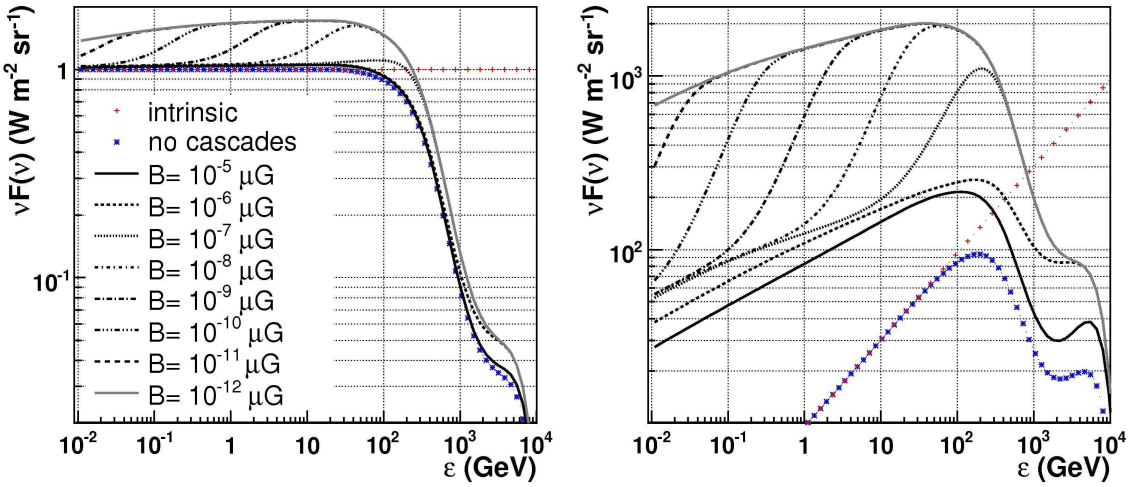

Figure 2.22: Calculated effect of the secondary (cascade) emission on the observed HE-VHE gamma-ray spectrum of a gamma-ray source at $z = 0.186$. The intrinsic spectrum is assumed to be a power law with spectral index $\alpha = 2$ (*left*) and $\alpha = 1.5$ (*right*), respectively, with a hard cut-off at 1000 TeV. Shown are intrinsic spectrum (*red* '+'), absorption corrected spectrum without cascade contribution (*blue* 'x'), and the spectra including cascades expected for various different strengths of the intergalactic magnetic field B (*black and gray lines*). The EBL model used is that of Primack, Somerville *et al.* (2001), but renormalised with a factor of 0.45. Adapted from d'Avezac (2006).

3. The EBL flux in the mid- to far-infrared band is not too low. The higher the EBL flux, the more above two requirements can be relaxed.

If these conditions are fulfilled, it can be expected that gamma-ray fluxes from blazars at intermediate to high redshifts (such as H 1426+428) are not only heavily attenuated above ~ 150 GeV by absorption in the EBL, but at the same time enhanced in the GeV range by a contribution from cascade gamma rays. This would have the net effect of increasing the gamma-ray horizon at GeV energies, by allowing the indirect detection of a certain fraction of those TeV gamma rays that have pair-produced in the EBL through the resulting cascades at the lower energies.

In as far as the interpretation of observed VHE gamma-ray spectra from extragalactic sources is concerned, it is near at hand to conclude that in this case:

- Disentangling the respective contributions of primary and secondary gamma-ray emission for sources at higher redshifts is not trivial, and requires a number of still uncertain input parameters (EBL-SED and strength of the IGMF) as well as a detailed simulation of the IC-scattering amplitudes.

- A simple "de-absorption" of observed spectra by application of (energy and redshift dependent) absorption factors derived from EBL models is no longer sufficient, and may lead to incorrect results.

Sensitive spectral measurements in the MeV-GeV part of blazar SEDs should allow for a positive identification of contributions from cascade emission, in particular when combined with improved gamma-ray emission models. New ground-based measurements at around 100 GeV threshold energy, and the upcoming GLAST gamma-ray mission with its unprecedented sensitivity in the MeV to low GeV range (see Fig. 2.21), should be able to provide a definite answer to the question of whether the observed gamma-ray spectra of blazars are measurably affected by a contribution from gamma-ray cascades or not.

CHAPTER 3

ATMOSPHERIC CHERENKOV TELESCOPES

After having introduced the basic concepts of cosmic gamma-ray detection in Chapter 1, we shall now examine the atmospheric Cherenkov technique in detail. The STACEE experiment, used to make the observations that form the basis of this work, is an atmospheric Cherenkov telescope. Just as important as an understanding of the detector itself, which will be described in Chapter 4, is a detailed knowledge of the physics of air showers and of the resulting Cherenkov light emission. This is because for ACT, the atmosphere, which acts as a converter and calorimeter, becomes effectively part of the detector. Providing the corresponding physics background for understanding the performance of these detectors is the goal of this chapter, followed by an introduction to the different detection methods utilised by current ACT.

3.1 Atmospheric Cherenkov light detection

As mentioned in Sec. 1.5, there currently exist three principal techniques by which high energy gamma rays from cosmic sources can be detected, each having its maximum efficiency in a different energy band. At the present time, experiments employing the atmospheric Cherenkov technique, *i.e.* detect the Cherenkov radiation produced by gamma-ray initiated extensive air showers (EAS), are producing the most exciting results within the field of gamma-ray astronomy (see also Sec. 1.5). The main reasons for this are the technique's capability to access a poorly explored energy band that contains a large number of physically interesting sources, and the fact that ACT are currently benefiting from rapid advancements in design and instrumentation. The technical improvements in ACT development result in continuously lowered energy thresholds and increased sensitivities, allowing for a progressively deeper penetration into the hitherto poorly explored energy range between 30 and 300 GeV. Even when the next generation of satellite-borne detectors, which will have sensitivity at higher energies than their predecessors, will begin to produce results, this is unlikely to change; satellite experiments remain limited by their relatively small collection areas, high costs and limited

upgradeability. In this section we will review the physical principles behind the emission of atmospheric Cherenkov light and its detection, as well as compare the properties of air showers from different kinds of primary particles.

3.1.1 Physics of Extensive Air Showers

The Earth's atmosphere is opaque to all electromagnetically interacting high-energy particles from extraterrestrial sources. Both cosmic rays, consisting of electrons, protons, helium and heavier nuclei (mainly C, N, O and Fe; see Sec. 1.3), and gamma rays are subjected to Coulomb interactions with air particles high in the atmosphere, initiating extensive air showers. The properties of the resulting air shower, characterised by the physical reactions through which it propagates, differ slightly depending on whether the primary particle was a gamma ray or electron on the one hand, or a hadron on the other hand.

Gamma-ray showers

At energies above 30 MeV, gamma rays lose energy in matter exclusively through pair production, dividing the initial energy approximately equally between the resulting electron and positron. For a given incident VHE gamma ray, the initial electron-positron pair is created after, on average, one radiation length (about 37 g cm^{-2} in air), translating into a first-interaction height of between 10 and 15 km ASL for vertical angle of incidence. As long as their energy is above a critical energy E_c, which is between 80 and 86 MeV in air (see Barnett *et al.*, 1996; Heck *et al.*, 1998), electrons and positrons lose energy predominantly via bremsstrahlung on air nuclei. This happens after roughly one radiation length, and reduces the electron's (positron's) energy by a factor of 1/e on average. The secondary gamma rays emitted in this process pair-produce again, giving rise to an *electromagnetic cascade*, driven by repeated pair-production and bremsstrahlung processes. Due to the high Lorentz factors of the particles in the cascade, the directionality of the initial gamma ray is conserved and the cascade propagates down through the atmosphere within a narrow cylinder centered around the extrapolated trajectory of the primary particle. This extension of the original gamma-ray trajectory is also called the *shower core*. The lateral distribution of charged shower particles scales with the Molière radius,

$$r_{mol} \;=\; 0.0212\,\mathrm{GeV} \;\cdot\; \frac{\xi_0}{E_c} \;\;=\; 9.6\ \mathrm{g\ cm^{-2}}/\rho_{air} \tag{3.1}$$

where ξ_0 is the radiation length, varying with the air density ρ_{air} and thus altitude, and E_c is the critical energy. Integrated over the range of contributing air densities, r_{mol}

turns out to be about 75 m. About 90% of the energy of an electromagnetic air shower is deposited inside a cylinder around the shower axis with radius r_{mol}.

Initially, the number of particles in the cascade increases exponentially. The further the air shower develops, the larger the fraction of electrons and positrons whose energy is less than E_c, the energy at which ionisation begins to dominate over bremsstrahlung as the principal energy loss mechanism. As a result, the number of particles in the cascade reaches a maximum and then declines rapidly, until the shower "dies out". The point at which the cascade has reached its maximum size, defined in terms of the number of bremsstrahlung emitting shower particles, is called the *shower maximum*. The location of shower maximum depends on the energy of the primary gamma ray: the higher the energy, the further into the atmosphere the shower penetrates, as the secondary particles have more energy to pass on to their daughter particles and therefore more pair-production/bremsstrahlung interactions can occur before the mean particle energy drops below E_c. For example, an air shower due to a vertical 10 GeV gamma ray reaches its maximum in average at 12.3 km ASL, while shower maximum for 1 TeV gamma rays is at around 8 km altitude ASL. Only cascades from gamma rays with energies well above 1 TeV penetrate deeply enough into the atmosphere for some of the shower particles to survive to sea level. This effect is illustrated in Figure 3.1, which shows the average longitudinal profiles of electromagnetic air showers for primaries of different energies. The atmospheric depth X_{max} of shower maximum for an electromagnetic shower can be calculated in a straightforward manner; one approximation of the classical cascade theory yields the expression (Konopelko, 2005)

$$X_{max} = X_0 \left(\ln \frac{E_0}{E_c} - \frac{1}{2} \right) \tag{3.2}$$

where E_0 is the energy of the primary gamma ray and X_0 the radiation length in air (37.1 g cm^{-2}). Using a realistic atmospheric model, an atmospheric depth X translates into height above sea level h for vertical showers given by

$$h = (6740\,\mathrm{m} + 250\,(\mathrm{g\ cm})^{-1}\,X)\,\ln(1030\ \mathrm{g\ cm^{-2}}/X) \tag{3.3}$$

Due to the uniform development of electromagnetic showers, shower-to-shower fluctuations are small, *i.e.* any two gamma rays of the same energy and with the same arrival direction produce very similar air showers. The fact that the radiation length is small compared to the thickness of the atmosphere causes gamma-ray showers to remain relatively compact (see Eq. 3.1).

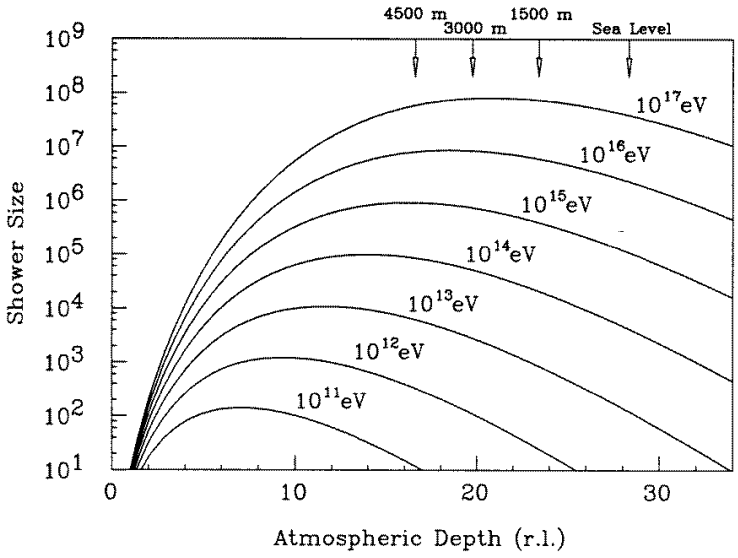

Figure 3.1: Longitudinal development of gamma-ray initiated air showers. Plotted is the average number of charged particles in the shower as a function of atmospheric depth, in units of radiation lengths (in air, one radiation length is about 37 g cm^{-2}). On the top, the corresponding altitudes above sea level for vertical showers are indicated. From Hoffman, Sinnis, Fleury & Punch (1999).

Hadronic showers

While the physics of electromagnetic showers is very well understood, the hadronic interactions taking place in air showers initiated by cosmic rays – henceforth called *hadronic air showers* – are, at the high energies (> GeV) concerned, still subject to large uncertainties. This is mainly due to the fact that energies obtainable by accelerator experiments on Earth are not sufficient to test Quantum-Chromodynamics (QCD) at those energies to sufficient accuracy. Due to the geometry of collider experiments, cross-sections in the GeV range and above are particularly poorly constrained for small centre-of-mass angles of the products with respect to the incident particles, *i.e.* in the longitudinal direction.

The fundamental difference between air showers initiated by hadrons and gamma rays is the fact that hadrons interact with the nuclei of air molecules not just electromagnetically, but also via the strong and weak nuclear forces. The qualitative description of a hadronic air shower is as follows. A hadron incident on the atmosphere will interact with air nuclei almost exclusively through production of mesons, which for the most part are pions. At primary energies above about a GeV, hadronic cosmic-ray nuclei disintegrate in the process, each fragment initiating a separate hadronic cascade. The neutral pions produced in these processes decay almost instantly (with a life time of, on

average, 10^{-16} s in their rest frames) to a pair of gamma rays, each of which initiates an electromagnetic sub-shower identical to the cascades resulting from primary gamma rays. Charged pions decay into muons after $2.6 \cdot 10^{-8}$ s in their rest frame, which – due to the time dilation at the high energies involved – is long enough to allow them to interact with air nuclei and produce more pions, giving rise to pionic cascades. While the resulting π^0 again create electromagnetic sub-showers, the charged pions driving the cascade eventually lose enough energy and decay to muons and muon neutrinos via the weak interaction:

$$\begin{aligned} \pi^+ &\longrightarrow \mu^+ + \nu_\mu \\ \pi^- &\longrightarrow \mu^- + \bar{\nu}_\mu \end{aligned}$$

Similarly, kaons and other heavier mesons may be created, which subsequently interact with air nuclei or decay into pions or muons. Several radiation lengths after the first interaction, a hadronic air shower will consist of a number of electromagnetic sub-showers with cores of hadronic particles.

As a result of the various interaction channels available to hadrons, there are several fundamental differences between hadronic and electromagnetic air showers. The larger transverse momenta available to the fragments from nucleonic collisions result in the air shower being more laterally spread out than a purely electromagnetic shower. Furthermore, because of the different particle compositions and four-momenta of the contributing sub-showers, a hadronic air shower will be less homogeneous than a purely electromagnetic cascade. As a further consequence and related to the above, fairly large differences between individual air showers are common, which are mainly driven by the nature and kinematics of the first interaction (and to a lesser extent by those of the subsequent hadronic collisions). Due to the pion decays, hadronic air showers are characterised by an abundance of muons which accompany the electromagnetic component. Due to the long live time of highly relativistic muons, and the fact that ionisation is the only significant energy loss mechanism affecting muons while traversing the atmosphere, muons will generally survive to ground level. They carry approximately 10% of a hadronic shower's energy. Schematic diagrams of the processes involved in electromagnetic and hadronic cascades is shown in Figure 3.2.

One notable fact about cosmic rays is that due to the initial disintegration of cosmic-ray nuclei in the atmosphere, air showers initiated by primary nuclei with atomic number Z and mass number A will have roughly the same properties as A superimposed proton

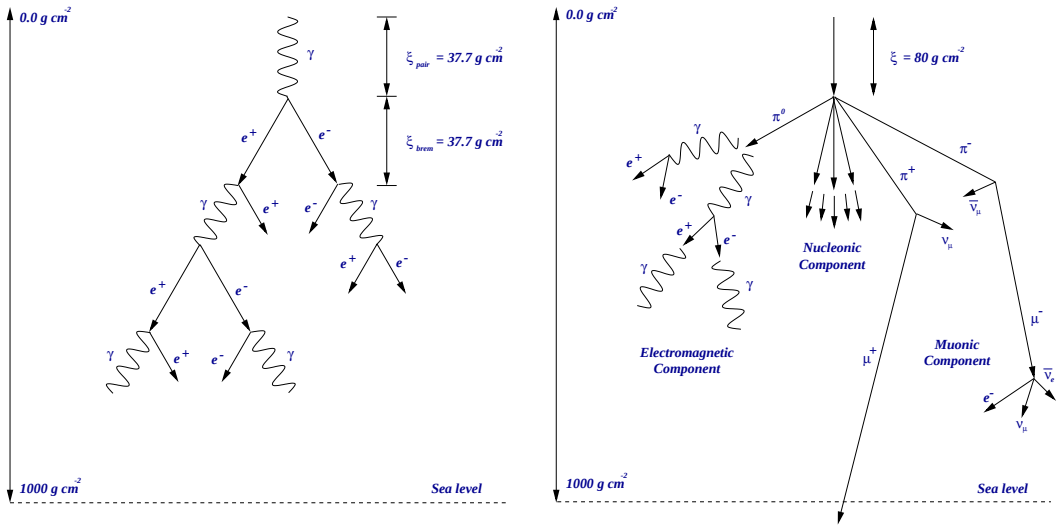

Figure 3.2: Sketch of the contributing processes in an electromagnetic cascade as created by highly energetic gamma rays or electrons penetrating into the atmosphere (*left*), and in a cascade caused by hadrons (*right*). Hadronic cascades contain nucleonic and muonic components as well as electromagnetic sub-showers from π^0 decays.

showers of energy E_0/A (with the caveat that the $(A - Z)$ neutrons do not interact electromagnetically).

3.1.2 Cherenkov light production

The physical principle allowing us to detect relatively low-energy gamma rays from the ground – and which is therefore utilised by most current ground-based gamma-ray detectors – is the so-called *atmospheric Cherenkov effect*. When a charged particle traverses a *dielectric* medium, *i.e.* a medium which is refractive and thus polarisable and transmissive to electromagnetic radiation, it causes a temporary polarisation of the ambient molecules. The depolarisation of the same molecules causes the emission of a faint radiation in the visible part of the spectrum, whose intensity however is normally much too low to be detected. The chief property of dielectric media is to slow down traversing electromagnetic radiation. In vacuum, photons travel with the speed of light c. In a medium with refractive index n (where $n > 1$), light propagates with the *phase velocity* c/n (here n is normally dependent on the photon wavelength λ). This means that in a dielectric medium, it is possible for highly energetic non-photonic particles to move *faster* than the local speed of light.

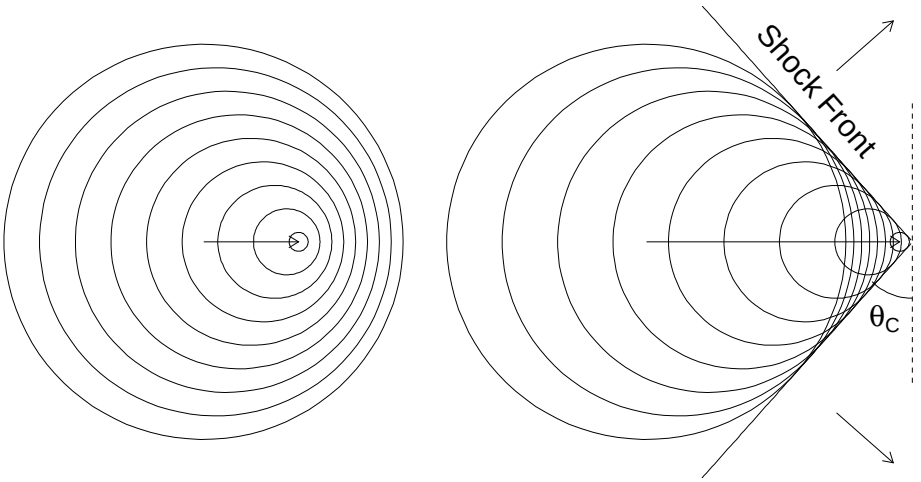

Figure 3.3: Principle of Cherenkov radiation emission. Faint blueish light from the depolarisation of air molecules is emitted spherically with phase velocity c/n around a relativistic charged particle traversing the atmosphere. *Left*: the particle is moving at half the speed of light in air ($0.5\ c/n$), and no shock front is created. *Right*: the particle's velocity is $1.33\ c/n$, causing a coherent superposition of the spherical wavefronts along a cone with opening angle θ_c. The resulting shock front constitutes the Cherenkov radiation. Note that the speed and Cherenkov opening angle (41°) shown here are typical for highly relativistic particles in water. In air, the opening angle of the Cherenkov cone is typically around 1°. Taken from Oser (2000).

If the velocity of a charged particle v is larger than the phase velocity $c/n(\lambda)$ in the medium that it is traversing, the emitted depolarisation radiation adds up coherently along a thin conical shock front propagating away from the particle trajectory at an angle θ_c, producing a brief flash of light along the shock front. This effect is principally the same as the sonic boom due to a projectile travelling faster than the speed of sound in air, and its geometry is illustrated in Figure 3.3. This phenomenon is called the *Cherenkov effect*, named after the Russian physicist Pavel Alekseyevich Cherenkov by whom it was discovered in 1934 while studying the effect of radioactivity on liquids.[1]

The *Cherenkov angle* θ_c, or opening angle of the Cherenkov cone, depends on the velocity $v = \beta c$ of the particle and the index of refraction $n(\lambda)$ of the medium, which itself is normally wavelength dependent:

$$\cos\theta_c \;=\; \frac{1}{\beta\, n(\lambda)} \tag{3.4}$$

[1]Pavel Cherenkov discovered only that the faint bluish light he observed was caused by β-radiation (electrons) and was independent of the target liquid, thus ruling out fluorescence. The actual explanation for the Cherenkov effect was found only a few years later, 1937, by his fellow countrymen Ilya Frank and Igor Tamm. In 1958, Cherenkov, Frank and Tamm shared the Physics Nobel prize for their work on this phenomenon.

It follows that in order for Cherenkov light to be emitted, the particle velocity $\beta = v/c$ has to be greater than $1/n(\lambda)$.

Air is a refractive medium with the property that $(n - 1)$ is approximately proportional to its density. In the Earth's atmosphere, air density decreases roughly exponentially with altitude. In extensive air showers, most of the Cherenkov light is emitted at shower maximum, since this is the point where the number of relativistic charged particles is largest. At this point, the Cherenkov production threshold according to Eq. 3.4 corresponds to several tens of MeV, and gradually decreases to about 22 MeV at ground level. Eq. 3.4 also states that the cosine of the Cherenkov angle is inversely proportional to the refractive index n of the medium. This means that due to the increase of air density with decreasing altitude, a focusing of the Cherenkov light takes place, resulting in a discrete ring on the ground where the photon density is highest. This focusing effect is illustrated in the right panel of Figure 3.4. In the limit of highly relativistic particles ($\beta \approx 1$) and assuming a constant refractive index n, the opening angle θ_c is thus given by $\arccos(1/n)$. For air, θ_c is around 0.7° at 10 km ASL (about the altitude of shower maximum), and approaches 1.34° at ground level (at ground level, $n = 1.000273$ for 20°C and 1 bar pressure).

The number of Cherenkov photons emitted per unit length $\mathrm{d}x$ and wavelength interval $\mathrm{d}\lambda$ along the particle's trajectory is given by

$$\frac{\mathrm{d}^2 N}{\mathrm{d}x\,\mathrm{d}\lambda} = \frac{(2\pi z e)^2}{hc\lambda^2}\left(1 - \frac{1}{\beta^2\, n^2(\lambda)}\right) \tag{3.5}$$

where ze is the particle's charge. The emitted photon flux as a function of the wavelength obeys an inverse square law, which in the case of air ($n \simeq 1.000273$) causes the majority of Cherenkov radiation to be emitted at blue to ultraviolet wavelengths. At wavelengths smaller than that, $n(\lambda)$ approaches unity and the right-hand factor of Eq. 3.5 causes a rapid decrease of the Cherenkov flux.

For a gamma-ray initiated shower, the radial Cherenkov photon distribution on the ground can be approximated by the empirical expression

$$\rho(R) \;=\; \frac{\rho_0}{1 + (R/R_0)^\beta} \tag{3.6}$$

Here $\rho(R)$ is the radial photon density at radius R from the shower core. ρ_0, R_0 and β are fit parameters which Konopelko (2005) found to be 14.0 m^{-1}, 264.0 m and 3.41, respectively, for a gamma-ray energy of 100 GeV and an observation height of 1800 m ASL. This fit function given in Eq. 3.6 provides a good description of the central part

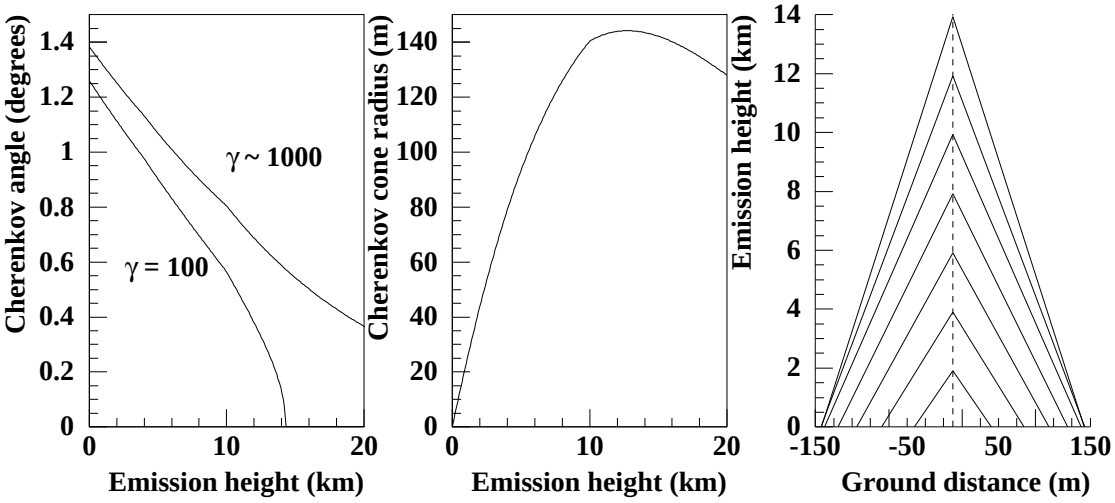

Figure 3.4: Geometrical properties of Cherenkov light emission in a vertical air shower. *Left*: Cherenkov angle θ_c as a function of emission height ASL, for particles with relativistic γ-factors of 100 and 1000 ($\gamma = E/mc^2$); for $\gamma \gtrsim 1000$, θ_c is approximately constant and depends only weakly on n. *Centre:* Radius of the Cherenkov ring at sea level as a function of emission height. *Right*: Diagram of the propagation of Cherenkov light as a function of emission height ASL, in the absence of multiple scattering, illustrating the focusing effect of the atmosphere which results in Cherenkov rings on the ground. From Fowler (2000).

of the Cherenkov light distribution, which is caused by the earlier arriving light due to the charged particles in the lower part of the cascade; however it fails to do so at radii of around 120 m to 130 m, where it does not well reproduce the narrow bump due to Cherenkov photons emitted high in the atmosphere. The energy dependence of the lateral density can be approximated by

$$\rho(E_0, R) \simeq \rho(R) \left(\frac{E_0}{100\mathrm{GeV}} \right)^{1.2} \tag{3.7}$$

The usefulness of the atmospheric Cherenkov technique is enhanced by the fact that the photon density on the ground scales almost linearly with primary energy E_0, allowing the atmosphere to be used as a giant calorimeter for incident gamma rays.

The *angular* distribution of the Cherenkov light is determined also by the lateral extent of the Cherenkov light emitting region and the given observation altitude. For a radially, (*i.e.* with impact parameter zero) incoming gamma-ray shower observed at sufficiently low altitudes, the angular size of the Cherenkov light emitting region is about 0.5°. This corresponds to the angular diameters of the full Moon and the Sun. The spatial distribution of arrival times of the Cherenkov light, also called the *wavefront*,

is roughly conical in shape. This is a result of the fact that the Cherenkov light emitting region of the air shower is elongated in longitudinal direction, and light produced later, *i.e.* in the lower part, will actually arrive on the ground *before* light emitted earlier at higher altitudes. As gamma-ray energies decrease, the longitudinal depth of the emission region shrinks, and the conical wavefront becomes more and more spherical. Thus, for gamma-ray energies below about 200 GeV, the Cherenkov wavefront can be well approximated by a sphere (see lower right panel in Figure 3.5).

3.1.3 Gamma-ray versus hadronic air showers

Quite fortunate for imaging Cherenkov telescopes is the fact that the lateral and angular distributions of the Cherenkov light at any given observation altitude are generally quite different for gamma-ray (and electron) initiated air showers on the one hand and hadron initiated air showers on the other. The homogeneous composition of purely electromagnetic showers gives rise to light pools with characteristics as described in Sec. 3.1.2: a bright ring centered at the shower core whose radius is determined by the observation altitude – and to a lesser extent the gamma-ray energy (for common altitudes and energies, it corresponds to about 130 m) – and an approximately uniform illumination inside that ring. Air showers due to hadronic primaries create very inhomogeneous Cherenkov light pools, caused by the various types of processes involved. The kinematics of the nucleonic interactions permit large relative momenta of the hadronic final states, which translates into large transverse momenta of the fragments (nuclei, nucleons or mesons). Most of these fragments ultimately end up producing neutral pions, which decay to two gamma rays. The vast majority of the Cherenkov light produced in a hadronic air shower stems from the electromagnetic sub-showers initiated by gamma rays from neutral pion decay, and therefore the dominating constituents of the Cherenkov light distribution of hadronic air showers are a number of laterally offset, superimposed circular "gamma-ray"-like Cherenkov light patterns. A certain fraction of the time, however, the first interaction of a hadron with an air nucleus transfers most of the primary particle's momentum to one single π^0, whose two daughter gamma rays then produce an air shower which is virtually indistinguishable from that of a primary gamma ray of the same energy. This fact sets a natural limit to the effectiveness of gamma-hadron separation methods. In hadronic air showers, isolated muons penetrating to ground level give rise to localised Cherenkov light ring patterns on the ground.

The differences in spatial distribution go hand in hand with different arrival time and angular distributions, which are correspondingly less homogeneous and more structured

for hadronic showers. In Figures 3.5 and 3.6, several properties of Cherenkov light distributions on the ground are plotted for a typical gamma-ray and proton initiated air shower, respectively. The radial profiles of both types of showers are compared in Figure 3.7. In hadronic showers, a large fraction of the primary particle's energy is distributed among hadronic particles, which due to their larger masses produce less Cherenkov light than electrons and positrons. Therefore, the total Cherenkov light yield of a hadronic shower is in average about a fourth that of a gamma-ray initiated shower of the same energy. The variation of the Cherenkov photon densities in dependence of primary particle type and energy is shown in Figure 3.8.

3.1.4 Cherenkov light detection

As mentioned previously, Cherenkov light produced in air showers arrives on the ground in the form of a brief (~ 5 ns long) flash of blueish to ultraviolet light. As the intensity of the Cherenkov flash is low (down to just a few photons per m^2 for 50-GeV gamma rays) when compared to the dominating isotropic background of night-sky light and, quite frequently, ambient man-made light, the key to its detection is the exploitation of the brief duration of the flash in order to discriminate it against this uniform background. To detect the Cherenkov light, sensitive photon detectors with nanosecond timing resolution are essential. The *photomultiplier tube* (PMT) is the most commonly[2] used light detector in gamma-ray astronomy, owing to its high gains of 10^5 to 10^6 and a good timing resolution. Another advantage of the PMT for the detection of Cherenkov light is the fact that its peak sensitivity lies in the UV to blue, whereas night sky background light peaks at longer wavelengths. PMTs use a series of *dynodes*, each at a successively higher positive voltage, to create a cascade of electrons proportional in size to number of primary electrons released (via the photoelectric effect) from a photocathode which is struck by the incoming photons. Photocathode and dynodes are usually made of an alkali metal with a low ionisation energy.

For a gamma-ray telescope, it is not enough to detect the Cherenkov light over the random fluctuations from night sky light. The Cherenkov light produced by hadronic cosmic-ray air showers constitutes an overwhelming background for gamma-ray detection. Unfortunately, due to the fact that ground-based gamma-ray telescopes only detect the primary particle indirectly by means of the resulting air shower particles, a complete or even just de facto elimination of this background is presently impossible. Instead, atmospheric Cherenkov telescopes rely on suppressing the cosmic-ray background as much

[2] At the time of writing, gamma-ray detectors use exclusively PMTs.

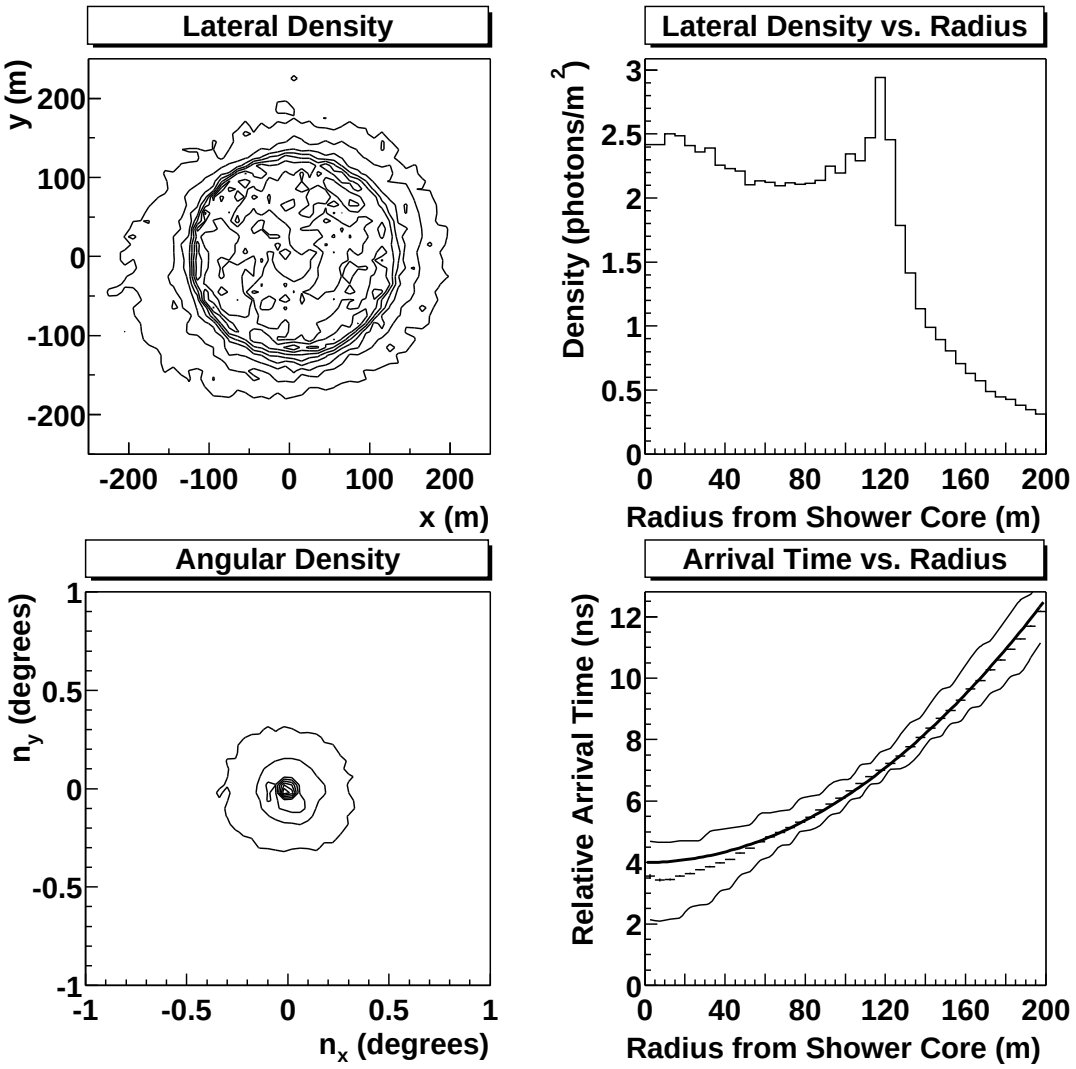

Figure 3.5: Properties of a single Cherenkov light flash due to an air shower from a vertically incident 100 GeV gamma ray, according to the *CORSIKA v5.946* air shower simulation. *Upper left*: Lateral Cherenkov light density on the ground relative to the shower core location. *Upper right*: Lateral Cherenkov light density on the ground as a function of distance from the shower core. *Lower left*: Angular Cherenkov light distribution in the plane normal to the shower arrival direction. *Lower right*: Mean Cherenkov photon arrival time at ground level (points with horizontal error bars) and the 80% contours (thin solid lines) as a function of distance from the shower core. The thick solid line is a fit to a perfectly spherical wavefront. Taken from Scalzo (2004).

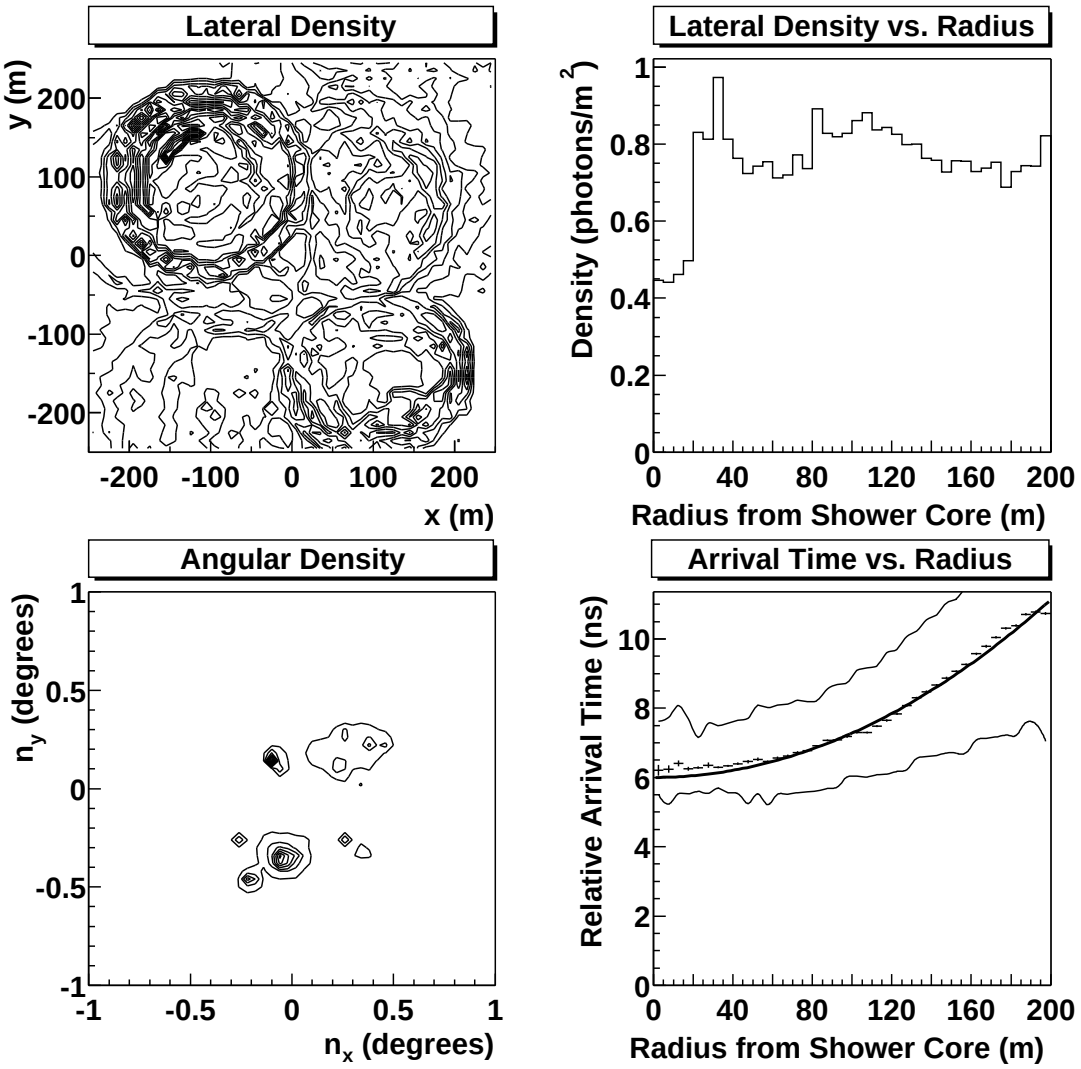

Figure 3.6: Properties of a single Cherenkov light flash due to an air shower from a vertically incident 400 GeV proton. according to the *CORSIKA v5.946* air shower simulation. *Upper left*: Lateral Cherenkov light density on the ground relative to the shower core location. *Upper right*: Lateral Cherenkov light density on the ground as a function of distance from the shower core. *Lower left*: Angular Cherenkov light distribution in the plane normal to the shower arrival direction. *Lower right*: Mean Cherenkov photon arrival time at ground level (points with horizontal error bars) and the 80% contours (thin solid lines) as a function of distance from the shower core. The thick solid line is a fit to a perfectly spherical wavefront. Taken from Scalzo (2004).

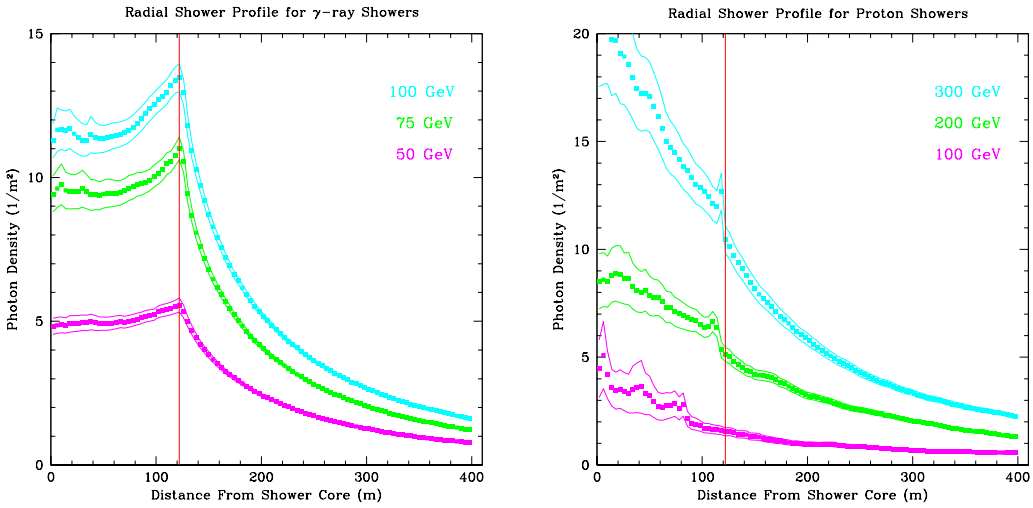

Figure 3.7: Typical Cherenkov photon density on the ground as a function of distance from the shower core, for various primary energies (gamma rays on the left, protons on the right), according to the *CORSIKA* air shower simulation (Heck *et al.*, 1998). Simulated showers are vertical and the observation level is set to 1700 m; the red line indicates the location of the gamma-ray Cherenkov ring for these parameters. Adopted from Boone (2002).

as possible, and quantifying it in a sufficiently accurate way. This is achieved through a combination of appropriate detector design, making use of Cherenkov light imaging and detection in cameras consisting of multiple pixels (*i.e.* light detectors) on one hand, and sensitive offline analysis techniques on the other. The former shall be examined in more detail in the following Sec. 3.2, while offline gamma-hadron separation techniques will be discussed in Sec. 3.2.1 for imaging Cherenkov telescopes, and in Sec. 6.6 for non-imaging detectors.

3.2 Detector types

In this section we will address the various existing concepts of atmospheric Cherenkov telescopes, differing mainly in the way in which the Cherenkov light is collected and focussed.

As remarked in Sec. 3.1.2, the total Cherenkov light yield – and thus the Cherenkov light density on the ground – scales almost linearly with the primary energy. Obviously, showers from higher energy gamma rays are easier to detect over a given background light level than showers from gamma rays of lower energy. On the other hand, due to the steeply falling spectra of all known gamma-ray sources and the existence of the poorly

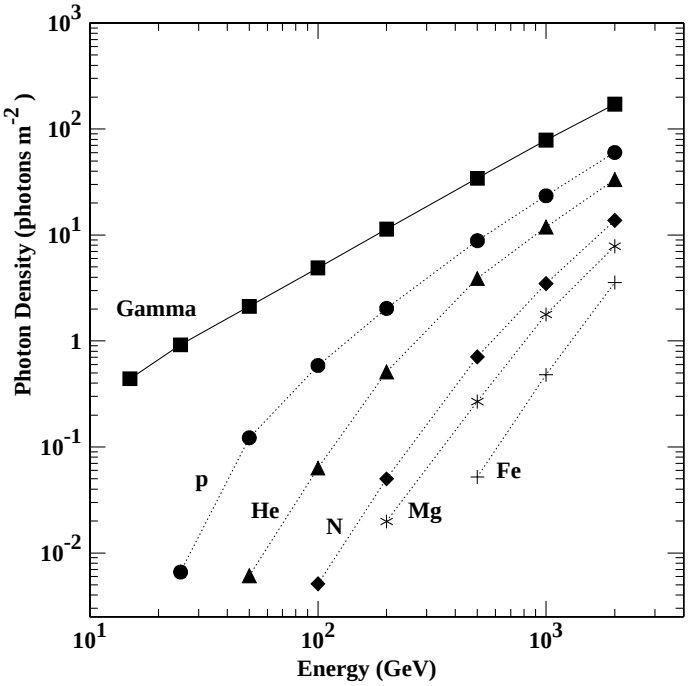

Figure 3.8: Average Cherenkov photon densities on the ground as a function of primary energy, for vertically incident gamma rays and various hadronic primaries, produced with the *MOCCA* air shower simulation (Hillas, 1985). The numbers shown are the densities of Cherenkov photons between 300 and 550 nm arriving at observation altitude (taken to be at an atmospheric depth of 830 g cm^{-2}, corresponding to 1900 m ASL), within a radius of 125 m of the shower axis and within 5 ns of the peak of the arrival time distribution. If a realistic field of view is simulated by applying angular cuts, the densities in the hadronic showers decrease by a factor of two to five. From Ong (1998).

explored energy window just above the energy range which can be efficiently observed with satellite-born instruments (see Sec. 1.5.2 and 1.5.4), the aim of any ground-based gamma-ray detector is therefore to extend its sensitivity towards as low an energy as possible, *i.e.* to "push down" the energy threshold as far as it is feasible. In general, this can be achieved in one of several ways.

In terms of the number of photoelectrons released from the photocathode of a PMT, the Cherenkov signal S obtained by a generic atmospheric Cherenkov detector can be written as

$$S \;=\; \varepsilon\, A\, \kappa(E)\, E \tag{3.8}$$

where A is the total collection area (*e.g.* of the mirrors), $\kappa(E) = \rho/E$ the (weakly energy dependent) ratio of the Cherenkov photon density ρ to the gamma-ray energy E, folded with the wavelength dependence of the PMT acceptance. ε is the combined average

efficiency of the photon detectors, *i.e.* the fraction of Cherenkov photons producing a photoelectron. The latter is commonly in the order of 10% and includes factors such as mirror reflectivity, the fraction of light collected by the PMTs, and the quantum efficiency of the PMTs. Similarly, the number of photoelectrons due to night-sky background B released within a given time interval τ – usually the time used for trigger formation – can be expressed as:

$$B = \varepsilon A \Omega \Phi \tau \tag{3.9}$$

Here Ω is the (average) effective solid angle viewed by a PMT, and Φ the background photon flux in units of $[\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{srad}^{-1}]$, folded with the PMT acceptance and integrated over wavelength. Combining Eqns. 3.8 and 3.9, one can express the signal-to-noise ratio of our prototype Cherenkov light detector as

$$\frac{S}{\sqrt{B}} = \kappa(E)\, E \sqrt{\frac{\varepsilon A}{\Phi \Omega \tau}} \tag{3.10}$$

If we define the energy threshold E_{thr} as some minimum signal-to-noise ratio required to trigger the experiment, it follows from Eq. 3.10 that

$$E_{thr} \sim \sqrt{\frac{\Phi \Omega \tau}{\varepsilon A}} \tag{3.11}$$

Eq. 3.11 provides us with a basic rule that applies to all types of ACTs, independent of their exact design. Decreasing the energy threshold of an ACT can thus be achieved by several means: a dark site, which will keep the NSB flux at a minimum; a good collection efficiency, determined by the optical design of the experiment as well as the choice of light-sensitive detectors; a small field-of-view and a short integration time; and a large collection area A. Most of these variables cannot be varied beyond a certain limit. This is obvious in the case of the night-sky background. The efficiency can be improved by using good optics (*e.g.* mirrors with high reflectivity) and high quantum efficiency PMTs, but the possible amount of improvement is clearly limited. Both solid angle and integration time have a hard lower limit determined by the spatial and temporal extent of the Cherenkov light shower; the field-of-view cannot be reduced below the actual size of the Cherenkov light producing region in the atmosphere (about 1° in diameter), and the time window cannot be made smaller than the duration of the Cherenkov wavefront ($\sim 5\,\mathrm{ns}$). The only quantity which can be – within practical limits – made arbitrarily large is the collection area A. For this reason, existing ACTs are being expanded by adding primary mirror area, and new designs of ACTs generally boast larger collection areas than existing detectors (see Sec. 1.5.3).

In our qualitative considerations above we only accounted for night-sky background noise, but not background in the form of Cherenkov light from showers due to hadronic primaries, or from secondary muons. While the latter type of background Cherenkov light can be readily eliminated by requiring a large enough area on the ground to be illuminated, discrimination against the former is not as straightforward. The aim is to reject hadron-initiated air showers by exploiting the intrinsic differences in the Cherenkov light patterns of the two types of air showers. In order to improve the ratio of detected gamma-ray to detected hadronic showers, a system of optical imaging devices is usually used, which assigns each of the individual light-sensitive detectors a designated, well-defined field of view. The spatial and temporal structure of the resulting pixellated signal can then be used to reconstruct some aspects of the air shower, which in an ideal case includes the primary particle type as well as its arrival direction and energy.

Conceptually there are two fundamentally different ways in which Cherenkov light from air showers can be imaged. These are described in the following.

3.2.1 Imaging Atmospheric Cherenkov Telescopes

Also called *imagers*, imaging ACT (IACT) record an actual image of the angular Cherenkov light distribution of an air shower. This is achieved with a concave dish-shaped optical reflector, usually consisting of many small spherical mirror facets, with a pixellated PMT camera at the focal point. For observations, the dish's symmetry axis is pointed at a direction sufficiently close to the sky region of interest, to ensure that it lies within the field of view of the reflector, and the selected point is tracked across the sky as the Earth rotates. The trigger rate due to cosmic-ray background is often estimated by altering the tracked sky coordinates slightly while the potential gamma-ray source is still in the field of view, usually alternating between two target positions sufficiently far away from the region of interest on either side of it.

The actual shapes of the reflectors vary, but the spherical Davies-Cotton design (Davies & Cotton, 1957) is currently the most common. Due to the unequal light travel times caused by the geometry of both spherical and parabolic reflectors, the width of the Cherenkov pulse is spread from 5 ns to about 8 - 15 ns at the focal plane of the reflector (Akhperjanian & Sahakian, 2004), depending on the size of the dish. Triggering is usually done by requiring the signals of several adjacent pixels to exceed a certain threshold within a coincidence gate long enough to account for the stretched pulses. Diagrams of two representative imaging telescopes, the prototype Whipple imager and one of the state-of-the-art H.E.S.S. telescopes, are shown in Figure 3.9.

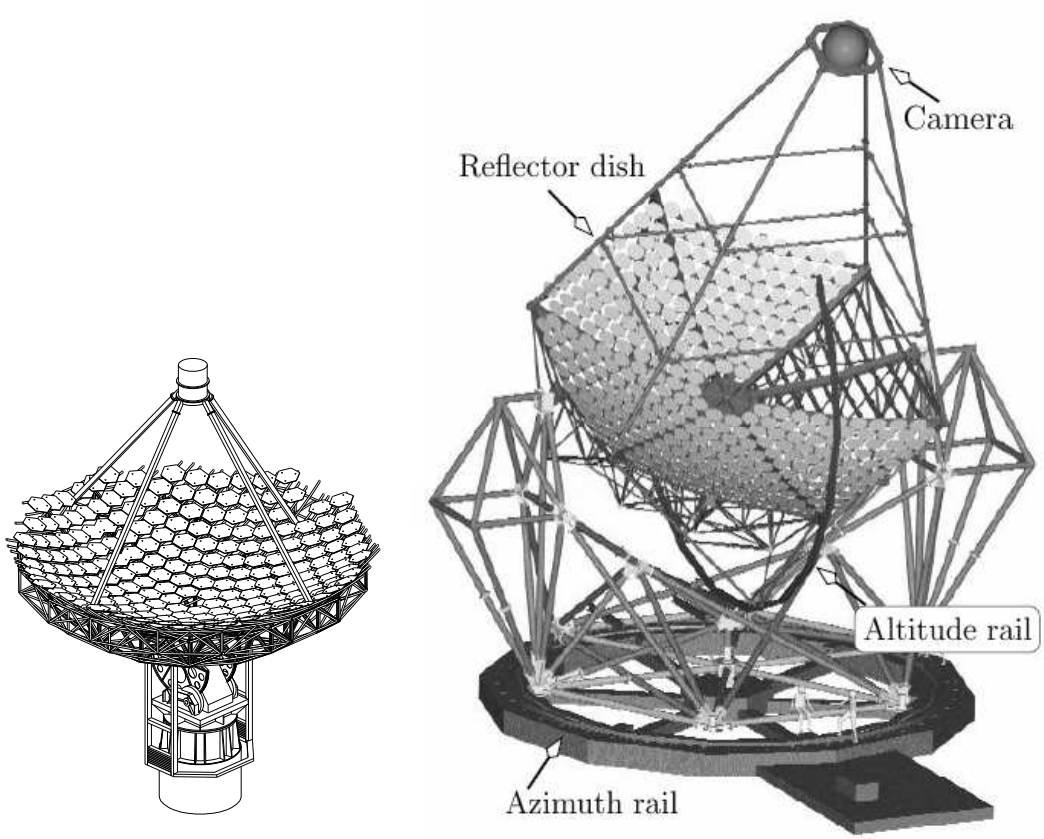

Figure 3.9: Diagram of the veteran 10 m diameter reflector of the Whipple observatory on Mt. Hopkins, Arizona (*left*, from Ong, 1998), and one of the four 12 m dishes of the H.E.S.S. telescope in Namibia (from Berge, 2002).

The raw image, consisting of different integrated pulse charges in a number of (PMT) pixels, is usually parameterized by fitting it to an ellipsoid. The reason for this is that the Cherenkov light emitting regions of gamma-ray showers appear as compact ellipses on the focal plane of the telescope, the degree of ellipticity and distance from the camera centre depending on the impact parameter of the primary (*i.e.* the shower core position). Images of hadronic showers generally exhibit large deviations from the elliptical template and are more irregular in shape. Figure 3.10 shows a typical example of a gamma-ray and a hadronic shower image recorded by one of the four HEGRA ACTs.

Figure 3.11 depicts the most important of the so-called *Hillas parameters*, providing a good first-order description of the shower in terms of the ellipsoid fitted to its image. This set of parameters was first used by the Whipple collaboration (Hillas, 1985; Hillas *et al.*, 1998). The strength of the imaging technique lies in the fact that it allows for a fairly accurate reconstruction of the arrival direction of the primary particle, which is the

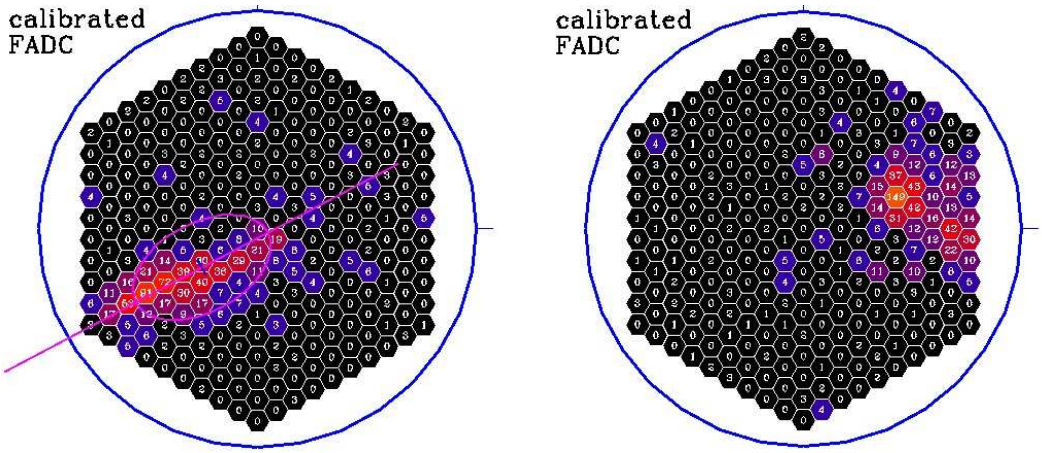

Figure 3.10: Images of a typical gamma-ray (*left*) and hadron (*right*) shower in the hexagonal camera of one of the HEGRA imaging ACTs. The number in each pixel stands for the corresponding height of the pulse within the coincidence window, in units of FADC counts (also indicated by the colour coding). In case of the gamma ray, the fitted ellipse is also shown. From www-HEGRA (2006).

major axis of the ellipse, and a fair estimate of the shower energy, obtained by integrating over the total charge of an image. By requiring a certain degree of ellipticity as well as cutting on a given arrival direction, it is also a powerful tool for rejecting isotropically arriving cosmic-ray showers. Using these basic but robust criteria, single imaging ACT are able to reject more than 99.7% of the hadronic background while keeping $\sim 50\%$ of the gamma-ray signal (Ong, 1998).

As mentioned in Sec. 1.5.3, all new ACT designs consist of arrays of imagers rather than one single telescope. Apart from achieving a larger mirror area and thus lower energy threshold, arrays of imaging telescopes allow shower reconstruction in stereo mode, which is illustrated in Figure 3.12. In the analysis, the images of all telescopes are superimposed, providing a far better angular resolution and rejection of hadronic showers than in single-telescope mode. Furthermore, single imaging telescopes are plagued by a considerable background of isolated muons, which produce a fair amount of Cherenkov light at very low altitudes which is imaged as a ring or section of a ring. While the rejection of complete or semi-complete rings is generally no problem for offline analysis algorithms, shorter segments of muon rings can easily be misidentified as gamma-ray images. The stereo imaging technique completely eliminates the background from isolated muons.

As the most sensitive and versatile technique for ACT, arrays of IACT will most likely be the technique of choice for all future Cherenkov telescopes.

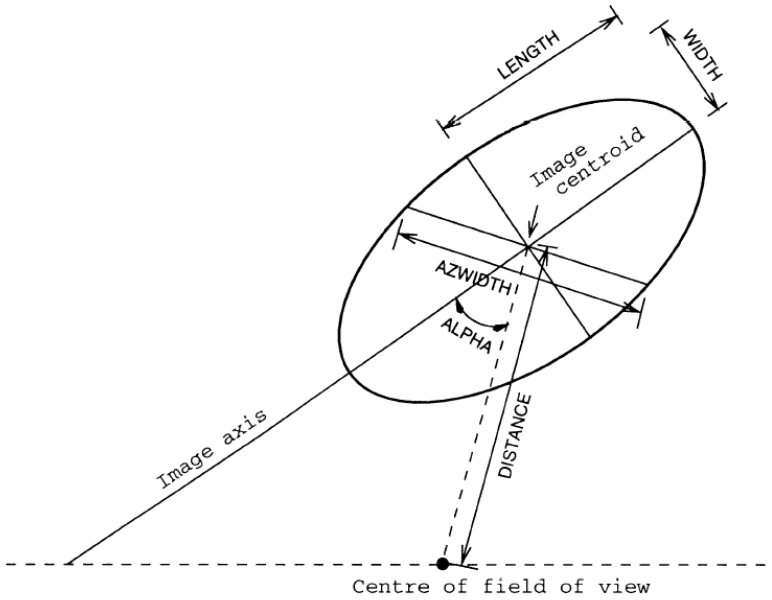

Figure 3.11: Image parameters used to characterise an IACT image, used for discriminating gamma-ray from hadron showers. All parameter distributions differ between gamma-ray and hadronic showers. Most important among these is the alpha parameter, which corresponds to the angle between telescope pointing direction and gamma-ray source for gamma-ray showers, but is uniformly distributed for showers from isotropic cosmic rays. From Ong (1998).

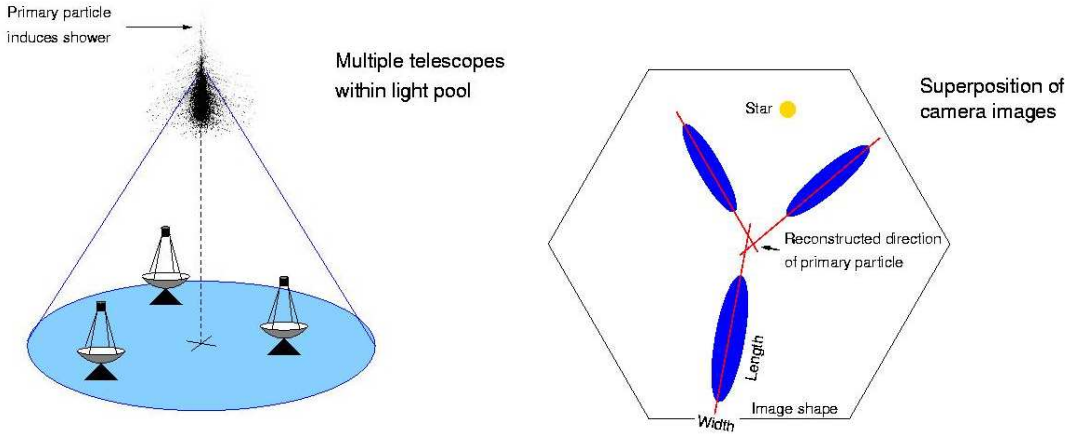

Figure 3.12: Principle of stereoscopic observations with arrays of two or more imaging ACTs. Individual images are superimposed and used to constrain the arrival direction of the primary particle, as well as to get a better estimate of the shower energy. Taken from www-HEGRA (2006).

3.2.2 Wavefront-Sampling Atmospheric Cherenkov Telescopes

A different approach to detecting atmospheric Cherenkov radiation is the so-called *wavefront-sampling* technique. Rather than using one or several telescopes with imaging capabilities, the aim of wavefront-sampling atmospheric Cherenkov experiments is to achieve a very large collection area by using arrays of large steerable mirrors to sample the Cherenkov wavefront on the ground at many positions simultaneously, each mirror viewing the whole of the Cherenkov light emitting region in the atmosphere. Current and past wavefront sampling telescopes have used existing – either operating or decommissioned – solar energy research facilities ("solar farms") for this purpose. This is mainly because constructing a dedicated wavefront sampling ACT from scratch would be very costly, but also most certainly because the principle of wavefront sampling has in the meantime been proven to be inferior to the imaging technique.[3]

The steerable mirrors of wavefront samplers require one or more stationary receiver facilities, usually housed on the top of a tower overlooking the mirror field, and secondary optics to focus the Cherenkov light onto cameras equipped with PMTs. The important difference between these types of experiments and imaging telescopes is the fact that "camera" pixels here are not mapped to different directions in the sky, but to different positions within the light pool of a Cherenkov light shower. Each pixel's field of view is large enough to contain the whole of the Cherenkov light emitting region, integrating over all possible directions. Thus wavefront samplers have no imaging capabilities (they are also called non-imaging ACT), but on the other hand the Cherenkov light received by each pixel is maximised. The drawback, or course, is that they have no resolving capabilities in angular space, making the rejection of hadronic background more difficult. Another consequence is that completely separate background observations are required to quantify the cosmic-ray trigger rate. These so-called "OFF" observations are normally performed immediately before or after the "ON" target observations, for the same duration and over the same range of local sky coordinates. This technique allows for a subtraction of the local background light, although as we will see in Sec. 6.3.2, it has some important restrictions.

Just as in the case of imaging ACT, wavefront samplers need to continuously update their pointing while tracking a source across the sky. In contrast to the former though, the geometry of the optics changes with changing target coordinates. The optimum observing position, corresponding to the largest effective mirror cross-section and thus optical throughput, depends on the actual design of the experiment but is usually located

[3] Albeit, a wavefront sampling ACT designed as such from the very beginning would in all likelihood be significantly more sensitive than any converted solar farms.

somewhere close to zenith. Any deviation from this position degrades the total optical throughput, and at larger zenith angles (elevations below $\sim 60^\circ$) the performance of most wavefront samplers is greatly restricted.

In the face of the aforementioned drawbacks, a good timing resolution ($\sim 1\,\mathrm{ns}$) is of the utmost importance for wavefront sampling experiments, in order to have any chance at overcoming the huge background of hadronic showers. Although unable to resolve angular space, sampling the Cherenkov light pool at many different positions on the ground and over a very large fraction of it allows the detector to measure the shape of the wavefront, provided that the timing of the detector elements is faster than the structure to be measured. Once the radius of curvature is found by recording the precise arrival times of the wavefront at the detector elements, in principle the primary direction can be found as long as a shower core position can be estimated.[4] The goodness of the fit of the arrival times to a curved wavefront in turn gives an estimate of how likely the shower originates from a gamma ray. This technique of discrimination against hadrons will be detailed in Sec. 6.6.1.

STACEE, which is a wavefront sampling detector, uses a two-level multiplicity trigger. The STACEE experiment will be described in detail in the following Chapter 4.

[4]In practice, it is problematic to lift the degeneracy between shower direction and shower core, especially at the lower gamma-ray energies of interest where the wavefronts are very close to spherical.

CHAPTER 4

THE STACEE DETECTOR

STACEE (Solar Tower Atmospheric Cherenkov Effect Experiment) was a large-area atmospheric Cherenkov detector located at the *National Solar Thermal Test Facility* (NSTTF) near Albuquerque, New Mexico, USA (34.96° N, 106.51° W), at an elevation of 1705 metres above sea level. Built in the mid-1970's, the NSTTF is a solar power research facility owned and operated by Sandia National Labs and the US Department of Energy (DOE). The facility includes an array of 222 large, independently steerable mirrors, called *heliostats*. These heliostats, each having a total mirror area of $37\,\mathrm{m}^2$, were designed to track the Sun and focus its light onto targets at various test bays on a 200' (61 m) tall *central receiver* tower (commonly referred to as *Solar tower*), located on the southern edge of the heliostat field. An aerial view of the facility can be seen in Fig. 4.1. During clear, moonless nights, the STACEE collaboration used 64 of these heliostats to collect Cherenkov light produced by gamma-ray induced air showers. The STACEE detector was operational from 1998 until 2007, in several different configurations (in terms of the number of used heliostats, signal readout, *etc.*).

The main motivation behind the construction of STACEE was the low expected gamma-ray energy threshold of a ground-based detector with such a large collection area, the idea being to bridge the energy gap between satellite-borne and ground-based gamma-ray telescopes (see Sec. 1.5.3). While imaging Cherenkov telescopes such as Whipple are able to map the angular distribution of the Cherenkov light emitting region and therefore have superior hadron rejection capabilities, their total mirror areas were up to that time restricted to the cross-sections of one or a few steerable spherical dishes. Out of this realisation, the idea of using the existing large mirror areas of solar farms was born.

The detection principle utilised by STACEE – and other atmospheric Cherenkov detectors using the wavefront sampling technique – is shown in Fig. 4.2. The Cherenkov light is reflected from the heliostats onto *secondary mirrors* located on two different levels of the Solar tower. The spherical secondary mirrors in turn focus the light onto arrays of *photomultiplier tubes* (PMTs), called "cameras", in such a way that each heliostat is mapped onto one single PMT. Copies of the PMT signals are then fed into a coincidence

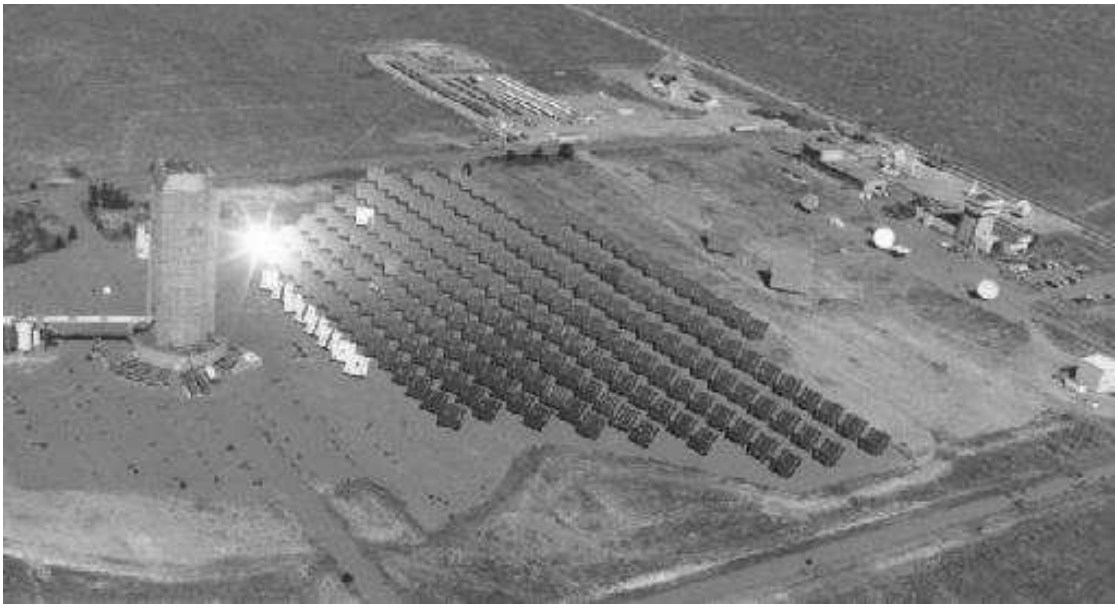

Figure 4.1: An aerial view of the National Solar Thermal Test Facility (NSTTF), at Sandia National Laboratories near Albuquerque, NM (USA).

logic in order to discriminate Cherenkov showers from random fluctuations in the night sky background, and upon satisfaction of a certain trigger condition, the PMT waveforms are read out, digitised and written to disk. The wavefront of the original Cherenkov shower can be reconstructed offline using the pulse height and relative timing information of the digitised traces.

In this chapter, a detailed description of the STACEE detector and its components is given. For a condensed technical review, see *e.g.* Gingrich (2005).

4.1 STACEE Optics

The motivation behind STACEE was to achieve – at least compared to imaging Cherenkov telescopes – a low energy threshold by taking advantage of the large mirror area of the existing solar array at the NSTTF. Naturally, such an endeavour is not without challenges. Trying to utilise a facility built for concentrating sunlight onto comparably large targets, without the need for a well focussed beam, necessitates the use of an appropriately designed secondary optical system that enables imaging of all chosen heliostats onto the individual "pixels" of the various PMT cameras. Furthermore, the acceptance of Cherenkov light from air showers of gamma rays with energies in the desired range should be made as large as possible, while at the same time the amount of accepted night-

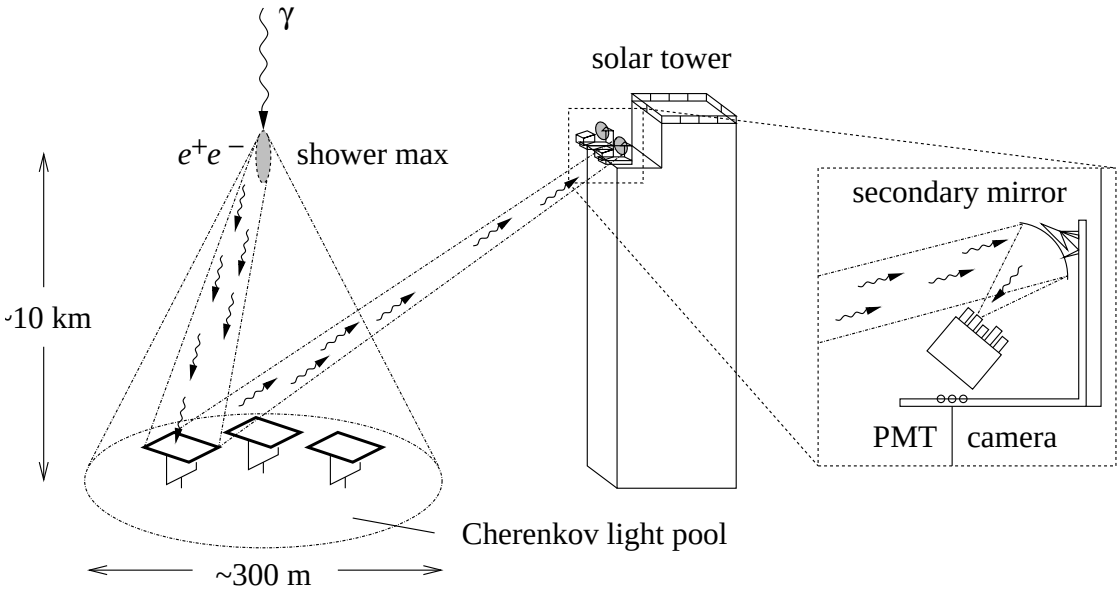

Figure 4.2: The concept behind STACEE. Atmospheric Cherenkov light from gamma-ray induced air showers is collected on the ground by large heliostats, and reflected to secondary mirrors on the upper levels of the central receiver tower. From there is focussed onto cameras of PMTs, so that each PMT receives light from only one heliostat.

sky background light (NSB) should be kept as low as possible (see Eq. 3.11). Finally, given the fact that atmospheric Cherenkov light is of blue to ultraviolet wavelengths, while the night-sky background light – and also the quite abundant ambient *man-made* background light – is dominated by longer wavelength components, the transmission of the optical system should ideally be optimised for light in the 300 - 450 nm range. In practicality these conditions can only partially be met, mainly because of the necessity to build on the existing facility.

In the following sections, the various components of the optical system will be described, along with the motivation behind their design.

4.1.1 Heliostats

Each STACEE heliostat consists of 25 individual square facets with a side length of 4 feet, mounted on a rigid metal framework in a 5×5 grid, yielding a total reflective area of 37.16 m^2 per heliostat. The facets consist of a layer of reflective silver between two protective layers of glass, and are held in place by five set screws, one in the centre and one on each corner. The optical alignment of each facet in the heliostat body frame can be changed through wrench adjustments on these facet mounting points. The same screws also bend the otherwise flat glass plates into approximately parabolic shape. During the

laser alignment procedure (described in Sec. 4.2.2), the focal lengths of the facets were adjusted to be roughly the distance from the heliostat to the top of the central tower. The framework holding the 25 mirror facets is attached to a fork-arm altitude-azimuth mount, whose orientation is controlled by two independent electric stepping motors, an *azimuth drive* at the base of the mount and an *elevation drive* attached at the end of the horizontal (elevation) axis. Fig. 4.3 shows a schematic diagram of a STACEE heliostat.

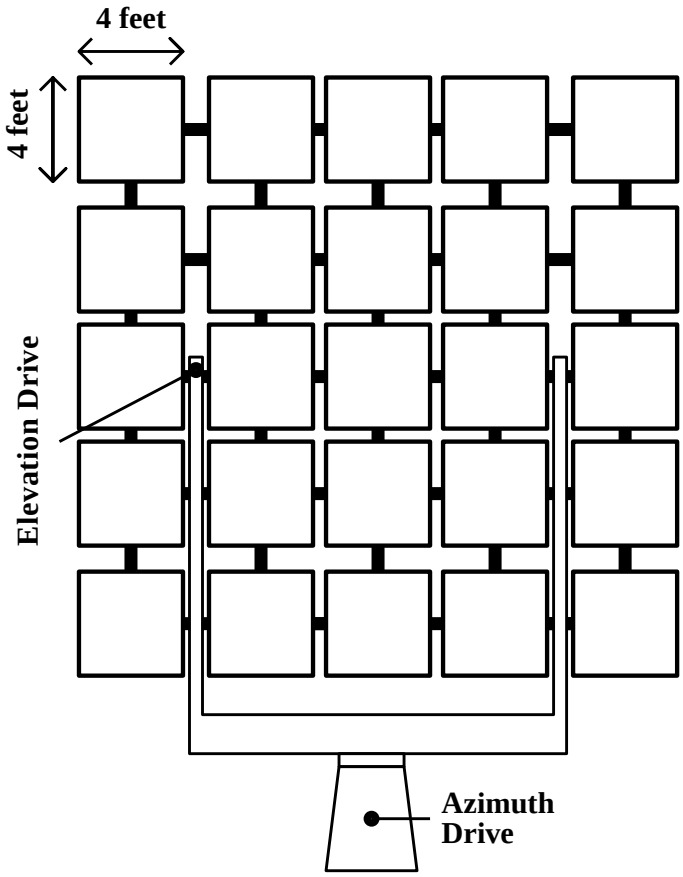

Figure 4.3: Schematic diagram of an NSTTF heliostat, showing the yoke and the 25 adjustable facets.

The reflectivity of the heliostat facets plateaus at close to 90% in the longer optical wavelengths, but due to the fact that light needs to traverse the front layer of glass two times, it is subjected to a cut-off in the ultraviolet, *i.e.* below 350 nm. The measured heliostat reflectivity as a function of wavelength is plotted in Fig. 4.17, along with the efficiencies of other STACEE components.

Each drive is equipped with a 13-bit position encoder, resulting in a pointing resolution of 0.044° per pointing bit in both elevation and azimuth. The drives have a fine and a course tracking speed. Fine tracking is used for tracking a source across the sky: the

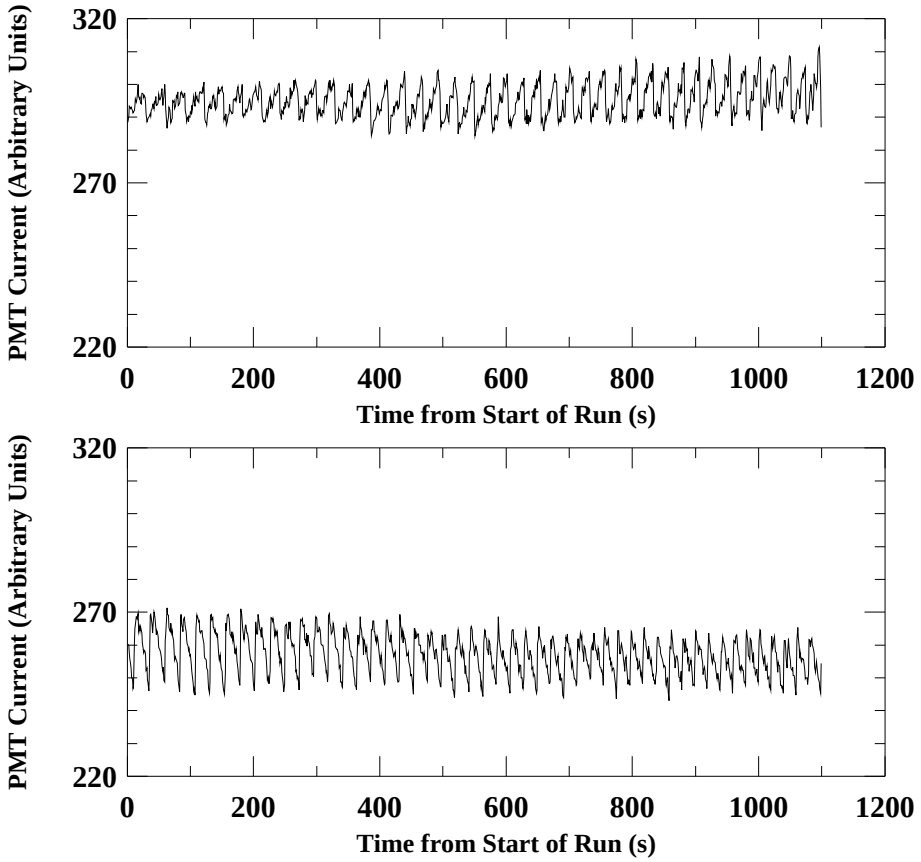

Figure 4.4: PMT current oscillations during star tracking as a result of the discrete pointing of the heliostats, which is updated whenever the difference between target and actual heading exceeds one pointing bit.

motors move the heliostat in discrete steps by individual bits whenever the accumulated offset between its heading and the desired pointing direction exceeds one encoder bit (0.044°). This happens *e.g.* when the sky position of the source has changed by more than 0.088° since the heliostat heading was last updated, which is about once every 20 seconds during normal tracking. The course tracking mode is used for *slewing* the heliostat to a different heading. Fig. 4.4 shows the phototube current due to the reflected light from a selected heliostat as a function of time. The oscillations are the result of a bright star in the heliostat's field of view and the fact that the heliostat pointing is updated in discrete steps.

For typical air showers from gamma rays with an energy in the range of interest to STACEE, the Cherenkov-light emitting region in the atmosphere is located at a height of about 11 - 13 km above sea level, and its angular diameter is about that of the full moon (0.5°). For this reason, the heliostats were not pointed parallel and right in the

direction of a prospective gamma-ray source during tracking, but rather at the point in the atmosphere where the bulk of the Cherenkov emission is expected to occur, *i.e.* at the shower maximum. This method of pointing the heliostats is called *canting* and is illustrated in Fig. 4.5. From now on we shall refer to this basic way of canting, in which all 64 heliostats are pointed at the expected shower maximum, as *monocanting*. From the 2003 - 2004 observing season on a different canting scheme, henceforth referred to as *paracanting*, was adopted. The paracanted configuration consisted of 48 heliostats canted as described above, and 16 interspersed heliostats pointed directly at the source position (also called *parallel viewing*). Data taken with the paracanted configuration allows for a better shower reconstruction, at the cost of a slightly elevated energy threshold (see Sec. 6.6.1).

The alignment and focusing of the heliostats were checked on a regular basis, using the procedures outlined in Sec. 4.2.1. The schematic in Fig. 4.6 shows the locations of the 64 STACEE heliostats with respect to the Solar tower, and their division into eight geographically connected 8-channel *subclusters*. The heliostats were chosen so as to ensure approximately uniform coverage in field area while at the same time avoiding an overcrowding of the photon detectors in the camera planes. The PMT cameras will be described in Sec. 4.1.3.

4.1.2 Secondary optics

In order to direct the Cherenkov light from the heliostats onto a camera consisting of individual photomultiplier tubes, suitable secondary optics were necessary. Because of the poor focussing of the heliostats and the very small PMT collection area, most of the Cherenkov light reflected to the camera by the heliostats would miss the PMTs, and light from one heliostat would "spill over" into many adjacent PMTs (called optical *cross-talk* in this context), if it were not collected and refocussed by a secondary optical device. For the reasons given in Sec. 3.2.2, the proper identification of Cherenkov light from gamma-ray showers required each heliostat to be mapped onto one single PMT channel; the individual images therefore needed to be condensed and separated before detection with a camera. Another purpose of the secondary optics was the restriction of the field of view of individual PMTs and thus the minimisation of the amount of accepted night-sky background light. In the case of direct viewing, the latter would be approximately equal to the cross-section of the associated heliostat divided by its distance from the PMT, which is strongly heliostat-dependent and generally in the order of a few degrees. When using secondary mirrors with appropriate diameters, the field of view is

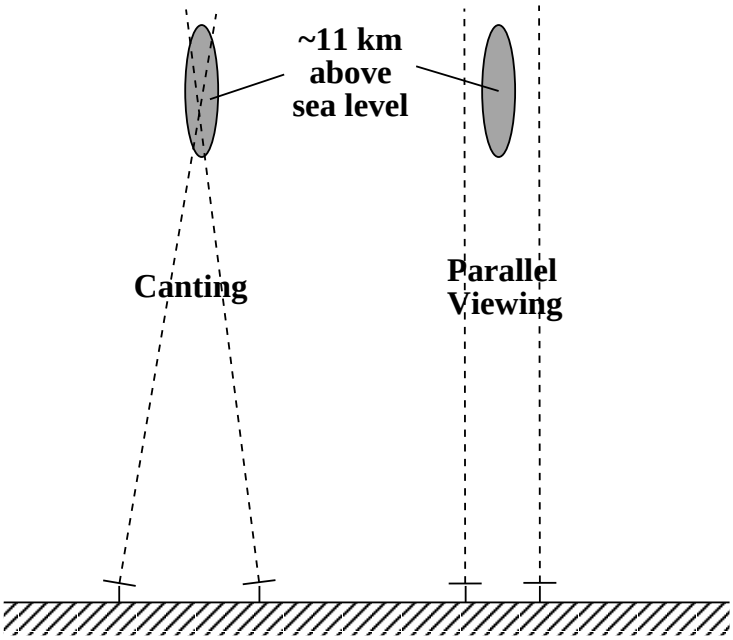

Figure 4.5: Sketch of the concept of canting. Instead of pointing directly at the source, all heliostats are viewing the expected interaction region in the atmosphere.

reduced to roughly half a degree, *i.e.* the typical size of the Cherenkov emitting region of a VHE particle-induced air shower. Optical light concentrators in front of the PMTs further reduced the individual channel's field of view. The PMT cameras are described in Sec. 4.1.3.

STACEE, as well as the CELESTE experiment in the French Pyrenees (Paré *et al.*, 2002), achieved this goal by using focussing mirrors on the tower which reflected the light from the heliostats onto the PMTs. In the case of STACEE, there was one secondary mirror for each PMT camera, which together with the support structure of the secondary (mirror) made up a *telescope*. These telescopes were installed on platforms at two different heights of the Solar tower: two mirror-camera units at 120 feet above ground (henceforth called the *220-level*, in referral to its height in feet above the foundation of the Solar tower) collected the light from the 16 STACEE heliostats closest to the tower, while the remaining 48 heliostats were viewed by three units at 160 feet above ground (on the *260-level*). Fig. 4.7 shows the locations of the STACEE secondary mirrors on two test bays of the Solar tower. A schematic view of a secondary mirror with PMT camera at the 260 level is shown in Fig. 4.8.

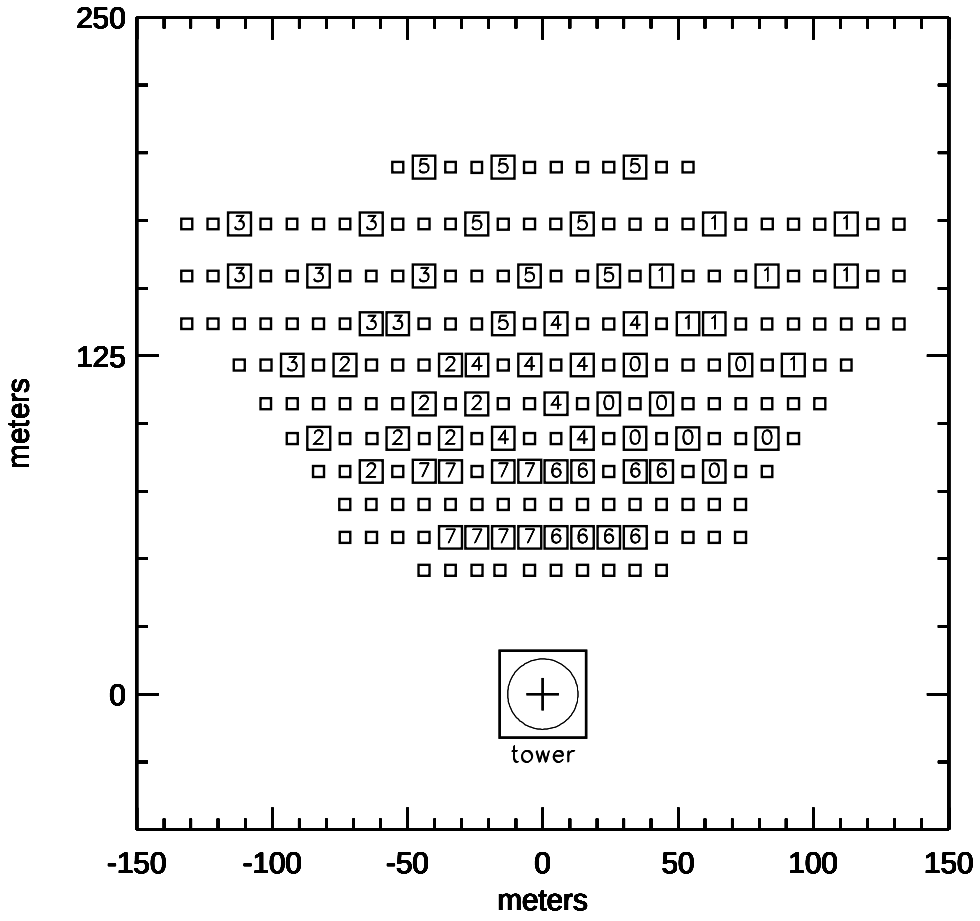

Figure 4.6: Map of the NSTTF heliostat field showing the heliostats selected for instrumentation as part of STACEE (North is up). The STACEE heliostats are numbered by the L1-trigger subcluster to which they belong (see Table 4.1). The mapping between PMT cameras and trigger subclusters is as follows: clusters 0 and 1 are viewed by the East camera, clusters 2 and 3 by the West camera, clusters 4 and 5 by the North camera, cluster 6 by the Southeast and cluster 7 by the Southwest camera.

All STACEE secondary mirrors were made of front-surfaced aluminised glass, which retains a high reflectivity of about 90% even at ultraviolet wavelengths (see Fig. 4.17), where most of the Cherenkov light from air showers is produced. The three secondary mirrors on the 260-level each consisted of seven hexagonal facets, which had a spherical surface and a focal length of 2.0 metres. Each of these composite secondaries had an average diameter 1.9 m and was mounted on an aluminum framework which could be raised into observing position and lowered for stowing after the end of observations by means of a hydraulic fork-lift. The headings of the six outer facets with respect to the

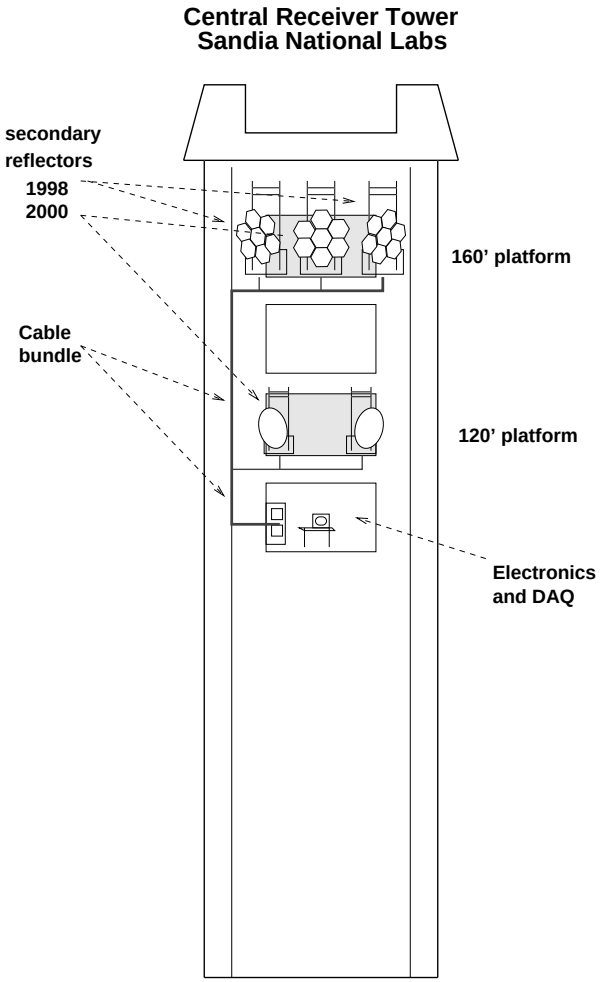

Figure 4.7: A view of the STACEE instruments at the Solar Tower.

fixed central facet could be adjusted via two bolts per facet, and were calibrated by reflecting a light source at the centre of curvature of the mirror (*i.e.* at a distance of 4 m from the mirror surface along the symmetry axis of the secondary mirror) onto a small placard at the same or slightly larger distance, and requiring that the individual images be co-aligned. The heading of the secondary mirror as a whole was controlled by means of two horizontal (one on each side) and one vertical actuator attached to the aluminum support frame. The secondaries were positioned so that the mirror axes described an angle of 32° with respect to the horizontal, "looking down" into the heliostat field. Neither the facet nor the secondary headings were adjusted during normal observing, but only when individual facets had been replaced or the optical alignment of the system

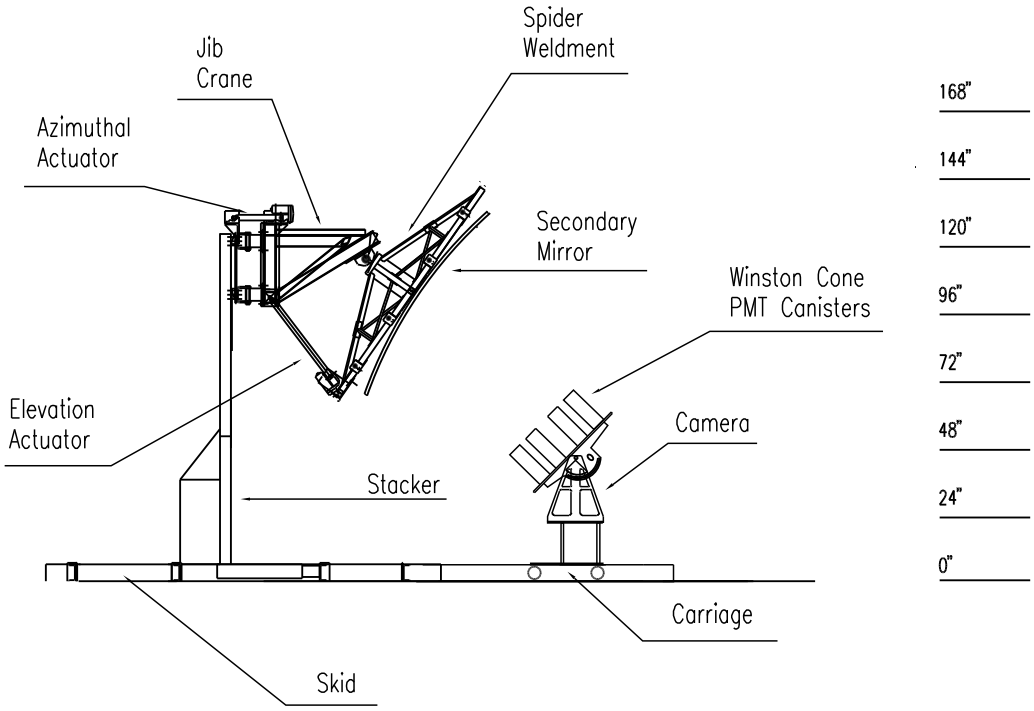

Figure 4.8: Schematic of secondary mirrors and cameras as used at the 260 level (East, West and North cameras).

was found to be inadequate. The positions of the 260-level secondaries – as well as the positions of the cameras – were monitored on a nightly basis to within half a centimetre by using a set of laser pointers and target grids installed on the camera plates and on the mirror mounts. The relative mirror–camera alignment had been found to be stable to approximately 1 cm, which corresponds to a change of about 5% in optical throughput. When not in use for extended periods of time, such as around full moon or in between observing seasons, the mirror facets were covered with protective plastic sheets. Photos of a 260-level secondary–camera system are shown in Fig. 4.9.

The two secondary mirrors on the 220-level were smaller and simpler in design. They were spherical as well, but consisted of one single round mirror facet 1.1 m in diameter, with a focal length of 1.1 m. Their headings could not be adjusted, rather they were mounted rigidly via a vertical support structure onto a sled which also supported the camera plates. The whole mirror-camera assembly slid back and forth on two I-beams and was deployed by being rolled out via a worm gear towards the edge of the platform before observing, and rolled back in afterwards. A garage door protected the detectors and all the equipment on the 220-level test bay from the effects of the weather when

Figure 4.9: Photos of the West (*left*) and East (*right*) PMT cameras with their secondary mirrors, located on the 260-level. The slotted camera plates with adjustable PMT mounts and the faceted secondary mirrors can be seen. (The photo on the right shows the camera in the state before the year 2000, without the box-like metal frame required by the weather protection tarps.) A STACEE collaborator is shown for scale.

the cameras were not deployed. Fig. 4.10 shows a photo of the Southeast camera and secondary on the 220-level.

4.1.3 PMT cameras

Each secondary mirror formed an image of the heliostat field at its focal plane. By placing individual light detectors at suitable locations at the focal plane, one can achieve a one-to-one mapping of heliostats and detectors ("pixels"). The design of the light detectors and the complete cameras are detailed in the following.

The STACEE *photomultiplier tubes* (PMTs) are described in Sec. 4.4.1. Each STACEE light detector consisted of a PMT and PMT base wrapped in copper tape (for magnetic and electrostatic shielding) and mounted inside a cylindrical steel canister of 11 cm diameter, also called a *can*. In order to collect the light from the image of a heliostat at the focal plane of the secondary mirror, which was typically a square with a side length of about 15 cm, and focus it onto the relatively small entrance aperture of the PMT (about 5 cm in diameter), a light concentrator was required. The concentrators used were so-called *DTIRCs* (Dielectric Total Internal Reflection Concentrators; Ning, Winston & O'Gallagher, 1987), which are conical transparent acrylic light collectors

Figure 4.10: Photo of the Southeast PMT camera with its secondary mirror, located on the 220-level of the Solar tower.

similar to *Winston* cones, but with slightly different geometry. The diameter of the entrance aperture was 11 cm, while the exit aperture size varied. The acrylic had an index of refraction of 1.49 and good transmissivity in the UV (see Fig. 4.17). Fig. 4.11 shows a photograph and ray-tracing diagram of a DTIRC. The well defined field of view of a DTIRC is determined by the opening angle of the cone and the diameter of the exit aperture. Light entering the front surface of the DTIRC from air at an angle less than a certain critical angle is totally internally reflected along the inside wall and exits through the exit aperture, whereas light incident at angles larger than the critical angle is lost through the DTIRC walls. Ideally, only photons passing through the exit aperture will hit the PMT photocathode and have a chance to be detected. The half-angle of the field of view of a DTIRC is therefore defined by this critical (or *acceptance*) angle. Fig. 4.12 depicts the photon density distribution at the exit aperture for a DTIRC with 24° acceptance angle, according to a ray-trace simulation.

STACEE used DTIRCs with three different acceptance angles – small (19°), medium (24°) and large (28°) – in order to partially equalise the different fields of view of the 64 STACEE PMTs. Since the DTIRC angle of acceptance determined the part of the

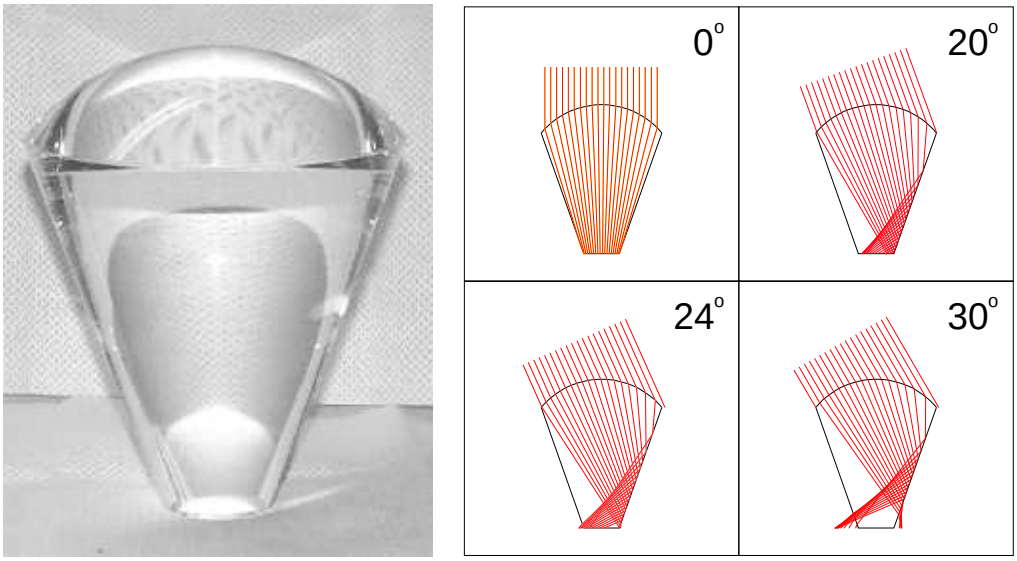

Figure 4.11: The DTIRC optical concentrators. *Left*: Photo of a typical DTIRC. *Right*: Ray-trace simulation of a DTIRC with 24° field of view, with parallel light incoming at various angles of incidence. Light incoming at angles less than 24° is internally reflected to the bottom aperture, whereas light incoming at angles greater than 24° is lost through the sides of the DTIRC. Taken from Oser (2000).

secondary mirror "seen" by the PMT, concentrators with small acceptance angles were chosen for the heliostats closest to the tower, which normally gave rise to the largest field of view (up to around 0.6° half-angle). Likewise, PMTs associated with more distant heliostats that had a small field of view (down to 0.25° half-angle) were equipped with large DTIRCs, permitting them to "see" almost the entire secondary mirror. PMTs viewing heliostats at intermediate distances were assigned medium DTIRCs. With the appropriate choice of DTIRCs, the viewing half-angles of all channels were equalised to between 0.26 and 0.37 degrees on the sky.

While it is common knowledge that the quantum efficiency of a PMT is dependent on the wavelength of the incident light, it is also a function of the exact incidence position and angle of the photon on the PMT photocathode. Therefore, one thing to take into account when using a light concentrator such as a DTIRC in front of a PMT is the fact that the angular and wavelength response of the combined system is not the same as the response of a "bare" PMT. As initially parallel light propagates through the DTIRC, the photons are subjected to a position-dependent deflection. As a result, the light density distribution at the DTIRC exit aperture, as well as the angular distribution of the photons exiting the DTIRC, will be non-uniform and strongly dependent on the original angle of incidence on the DTIRC. Furthermore, Snell's law causes the angular

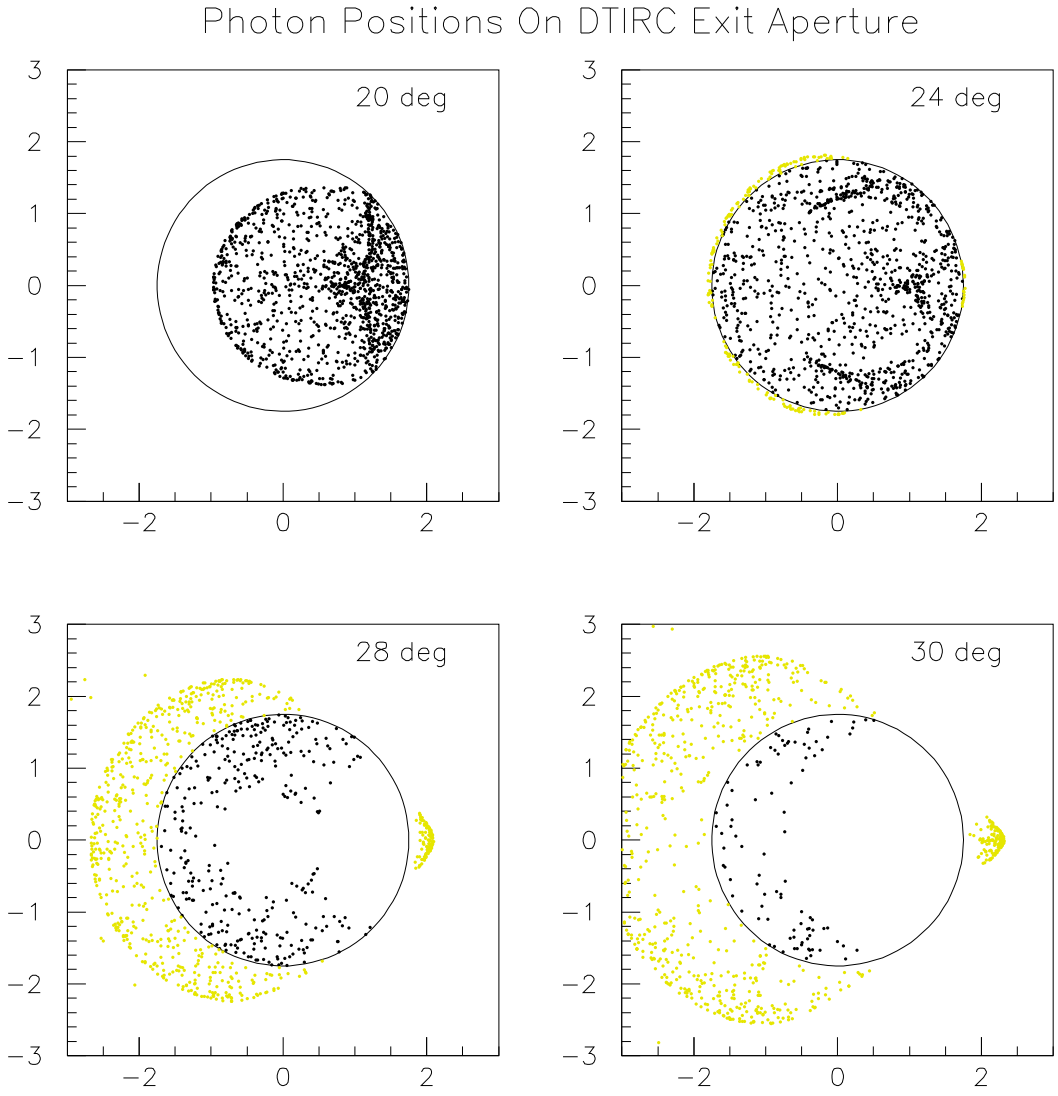

Figure 4.12: Photon density distribution at the exit aperture of a medium size (24° field of view) DTIRC, for various angles of incidence, according to a ray-trace simulation. Photons that exit through the side walls are depicted in yellow (light gray). For angles up to the critical angle (24°), all photons are able reach the exit aperture; for larger angles, most photons escape through the sides of the cone.

distribution of the photons to become even more spread out as the light transits from the dielectric of the DTIRC into the vacuum inside the PMT. Thus, for angles of incidence close to the critical angle, a large fraction of the light hits the PMT photocathode at shallow angles, giving rise to a net increase of the quantum efficiency. This effect is called *enhancement*.[1] In order to account for the enhancement, the responses of all cans used in the experiment were recorded using a test bench. This crude but effective device, built at the University of Chicago, consisted of 50 ft. long light collimation pipe with a white light source on one end and a turntable to hold the can on the other. The PMT response was mapped out by recording the PMT anode current as a function of the angle between can and incident light. Various filters (neutral density and narrow band interference filters with transmission at 400, 450, 550 and 600 nm) were placed in front of the light source in order to measure the wavelength dependence of the PMT response. An initial measurement of the relative magnitudes of the enhancement peaks was used to determine which size of DTIRC each PMT was eventually combined with: tubes with large enhancements were assigned to small DTIRCs and vice versa, in an attempt to equalise their angular responses. A measurement performed on the completely assembled cans was then used to model the response of the cans in the optical part of the detector simulation (see Sec. 5.3). Fig. 4.13 shows plots of typical PMT response curves as measured on the test bench.

The DTIRC was held in place by a steel retaining ring on the upper end of the can, and by the PMT and a ring-shaped plastic DTIRC holder from inside the can. The exit aperture of the DTIRC was optically coupled to the PMT entrance window via a 3.2 mm thin transparent silicon disk, commonly referred to as a *cookie*. On top of the cookie and surrounding the exit end of the DTIRC sat a thin black plastic ring also called a *donut*, whose purpose it was to prevent stray light from outside the DTIRC from entering the part of the circular PMT entrance window which was not in contact with the exit aperture of the DTIRC. The PMT was attached to the threaded base plate of the can via a long Allen screw and a circular piece of foam rubber (*pressure pad*), which was used to create enough pressure on the cookie to ensure good optical coupling between the DTIRC and the PMT entrance window, independent of ambient temperature. Both pressure pad and base plate had two circular openings to accommodate the signal and high-voltage (HV) connectors of the PMT. An exploded view of the complete can assembly is shown in Fig. 4.14.

[1]The enhancement is particularly strong for light of longer wavelengths (yellow, red). At these wavelengths, many PMTs exhibit enhancements of around 50% for DTIRC entrance angles around 20°.

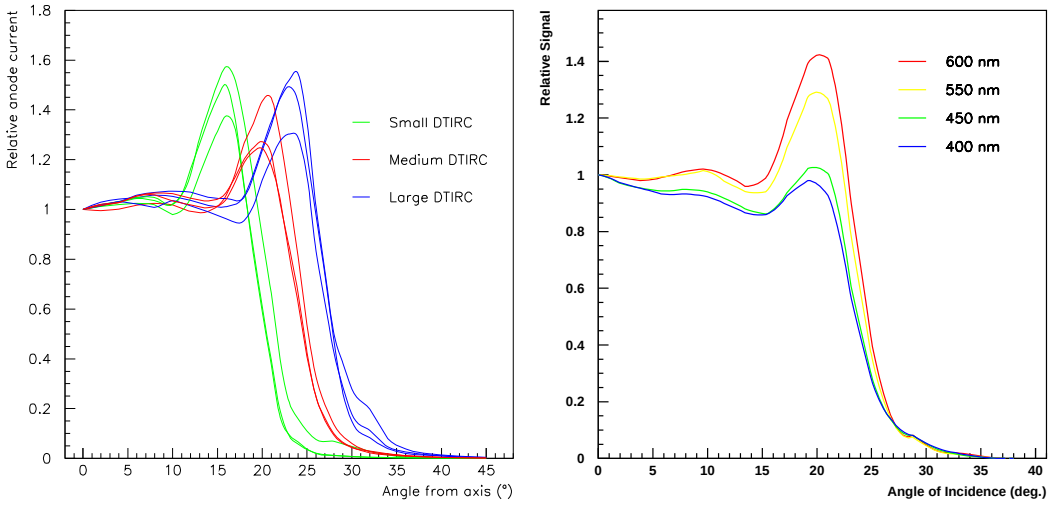

Figure 4.13: Measured response of DTIRC-PMT can assemblies as a function of the photon angle of incidence, in arbitrary units normalised to normal incidence. *Left*: Angular dependence of the response to white light from an Oriol lamp for nine representative cans, three of them equipped with small, medium and large light collector cones (DTIRCs) each. It can be seen that the magnitude of the enhancement, which is the peak of the response curve occurring at angles of incidence close to the cut-off angle, differs significantly between individual cans. *Right*: Angular dependence of the response for several different wavelengths for the example of a DTIRC with medium field of view (nominal cut-off 24°). From Fortin (2000) and Hinton (2000b).

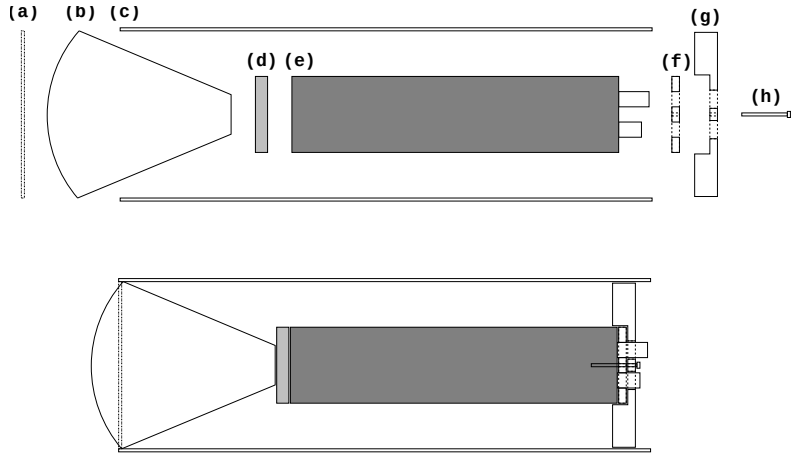

Figure 4.14: The standard STACEE PMT "can" assembly, in exploded (*above*) and assembled (*below*) views. Parts as labelled are: (a) retaining ring, (b) DTIRC, (c) can, (d) cookie, (e) PMT with PMT base, (f) pressure pad, (g) can base plate, (h) long Allen screw. Not shown: DTIRC holder.

The design of the PMT cameras included a metal plate with parallel slots in which the PMT cans were mounted via swivelling can holder bases, in such a way that their positions in the plane of the camera plate, as well as their headings, could be adjusted continuously. Furthermore, the exact vertical positions of the cans with respect to the camera plates were fine-tuned by means of plastic (ABS) cylinders of different lengths located in the can bases, on which the PMT cans sat. This was necessary to ensure that the front surfaces of all DTIRCs were level at the focal plane of the respective secondary mirror, independent of their position in the camera and their orientation. The camera plate dimensions were 1.7×1.0 m^2 for the 260-level and 0.8×0.5 m^2 for the 220-level cameras. The vertical tilt angle of the plates was adjustable within a certain range, but for proper operation the normals to the camera plates had an angle of 12.8° with the optical axis of the corresponding secondary mirror, meaning that the cameras were viewing the heliostats *off-axis* (see Fig. 4.15). This geometry had the advantage that the occultation caused by the large camera plates was minimised (albeit still significant for the PMTs closer to the optical axis, *i.e.* those viewing the closer heliostats). The disadvantage of the off-axis design as the fact that it increased the amount of spherical aberration, especially for heliostats far from the tower. Fig. 4.16 shows CCD images of the light distributions of selected heliostats; the "coma tails" due to aberration effects are clearly visible on some of the images.

Most of the hardware for the PMT cameras, including the camera boxes and DTIRC light concentrators, were designed and built at McGill University.

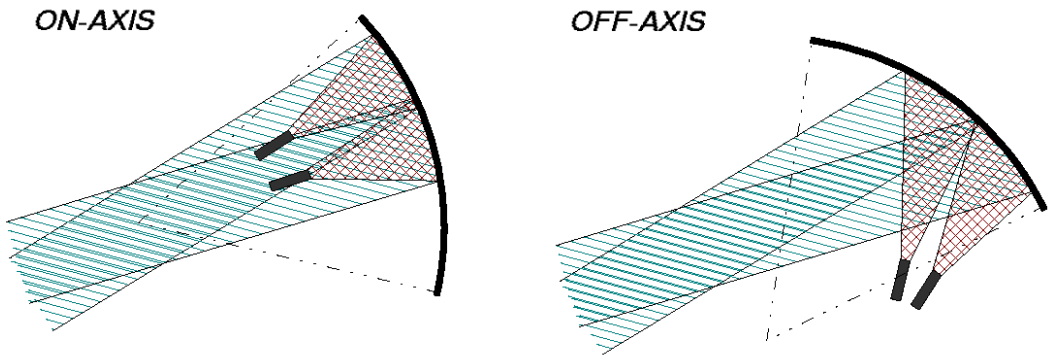

Figure 4.15: Different ways of camera placement with respect to a focusing mirror: the secondary optics of CELESTE and Solar-II (as well as all imaging ACTs) followed the *on-axis* design (*left*), which minimises spherical aberration. For STACEE, due to the large size of the camera plates, an *off-axis* design (*right*), which minimises occultation due to the camera, seemed favourable. Schematic taken from de Naurois (2000).

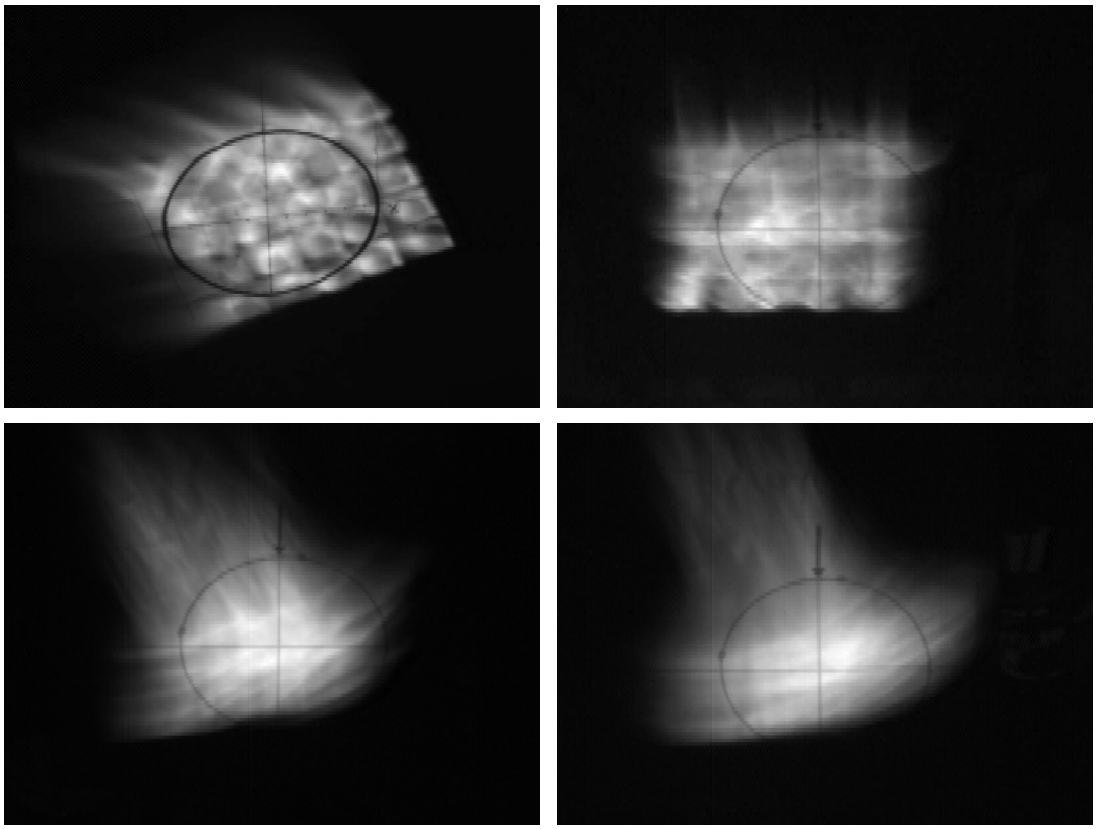

Figure 4.16: CCD images of camera spots, showing the light distribution and the target on the placard. The circle (appearing as an ellipse due to projection effects) marks the aperture of the DTIRC. The associated heliostats, in order of increasing distance from the tower, are 5W1 (*top left*), 8E6 (*top right*), 10E10 (*bottom left*) and 12E12 (*bottom right*). The coma tails due to aberration and the images of individual facets are clearly visible.

4.2 Optical alignment

The optical components of a wavefront sampling Cherenkov telescope are the most fundamental elements of the detector. Due to the high degree of sensitivity of the optical throughput to misalignments of the various elements (heliostats, secondaries and light detectors), and the need to minimise the amount of accepted background light, optimising the performance of the optical system with the help of proper alignment procedures is of utmost importance. At STACEE, maintenance or accidental mechanical disturbance of optical components, as well as replacements of parts such as mirror facets, heliostat drive motors or parts of the camera or secondary support structures, necessitated optical realignment of the affected elements. In addition, periodic checks of the alignment were performed in order to verify that the system was in the optimum state. In the

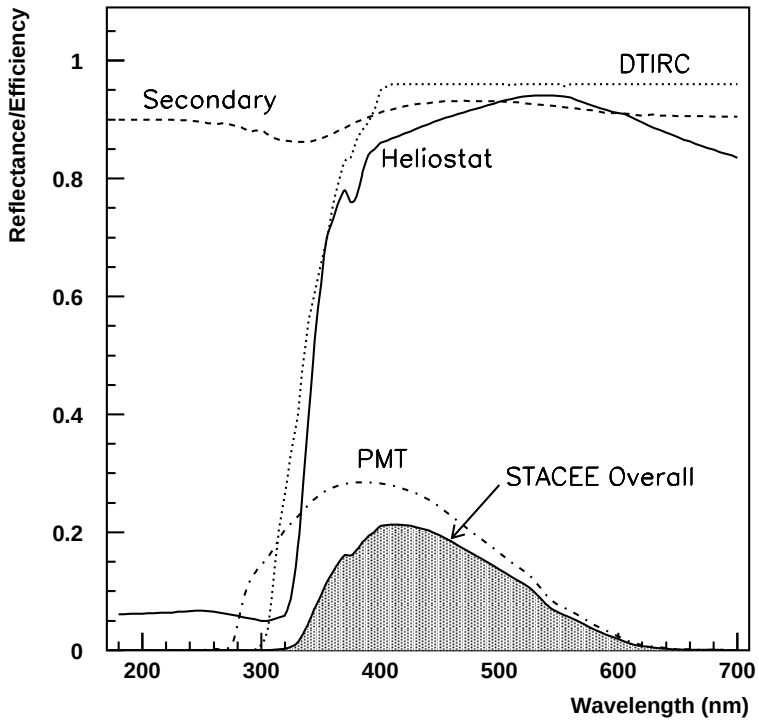

Figure 4.17: Efficiency curves for various optical elements in the STACEE experiment as a function of wavelength, including heliostat and secondary reflectivities, DTIRC transmission, and PMT quantum efficiency. The product of these curves (labelled "overall") is also shown.

following section, the various methods used to align the STACEE optical components are described in some detail.

4.2.1 Heliostat biasing

As mentioned in Sec. 4.1.1, the heliostats at the NSTTF utilise azimuth-elevation mounts, and both the azimuth and the elevation drives are equipped with position encoders of 13 bit resolution, corresponding to 8192 pointing positions per 360°. The mapping between encoder reading and actual heliostat heading needs to be determined by means of an appropriate calibration procedure. This procedure is called *heliostat biasing*. Several biasing methods have been adopted by STACEE over the course of its lifetime, but for the last few years of its lifetime only one of these, the *drift scan* method, was used to obtain the biases for astronomical observations. Nevertheless, a brief description of each method is provided in the following.

Drift Scans

The drift scan technique made use of any sufficiently bright star (magnitude m $\lesssim$ 3). Preferred were hot A-type stars due to the peak of their emission spectrum being at blue or ultraviolet wavelengths. The target objects were furthermore required to have local sky coordinates at the time of the scans that were within the range typical for Cherenkov observations, *i.e.* sufficiently close to the transit point and with a zenith angle of less than about 25°. The reason behind this was two-fold. Firstly, it had been found that the absolute pointing direction of a heliostat was not quite linear with respect to the set azimuth and elevation encoder bits, due to imperfections of the heliostat mount (see Sec. 4.2.3 for details). This gave rise to a direction dependent offset from the nominal pointing direction, which became larger with departure from the heading used to set the heliostat biases. Secondly, aberration effects due to the off-axis orientation of the heliostat facets, amplified by the secondary mirror optics, increased with increasing difference between source position and the tune-up heading used during the laser alignment procedure (azimuth 180°, elevation 70°; see Sec. 4.2.2 for details).

At first, all heliostats were made to track a point in the sky which was a certain angular distance (typically one degree, *i.e.* 4 minutes in RA) West of the star, but at the same declination. The heliostats were then commanded to stop tracking, allowing the star to drift through their fields of view. At the same time the PMT anode currents were recorded as a function of time. This procedure was repeated several times for tracking positions offset in declination with respect to that of the star, by various amounts in either direction. In this way, a two-dimensional "current map" was obtained for each heliostat. The locations of the peaks of these distributions, caused by the light from the star, were then used to calculate the differences between commanded and actual pointing direction of the heliostats in terms of azimuth and elevation encoder bits. Fig. 4.18 illustrates the peak-finding procedure.

The advantage of this method lay in the fact that it does not rely on a particular celestial object such as the Sun or Moon, and that data on all heliostats could be taken simultaneously. Furthermore, special filters made of 1-mm thin *BG3* glass that were placed in front of the DTIRC entrance apertures extended the range of possible target stars towards higher brightness, while at the same time suppressing the background light. Using drift scans, the heliostat pointing could be set to an accuracy of about one pointing bit.

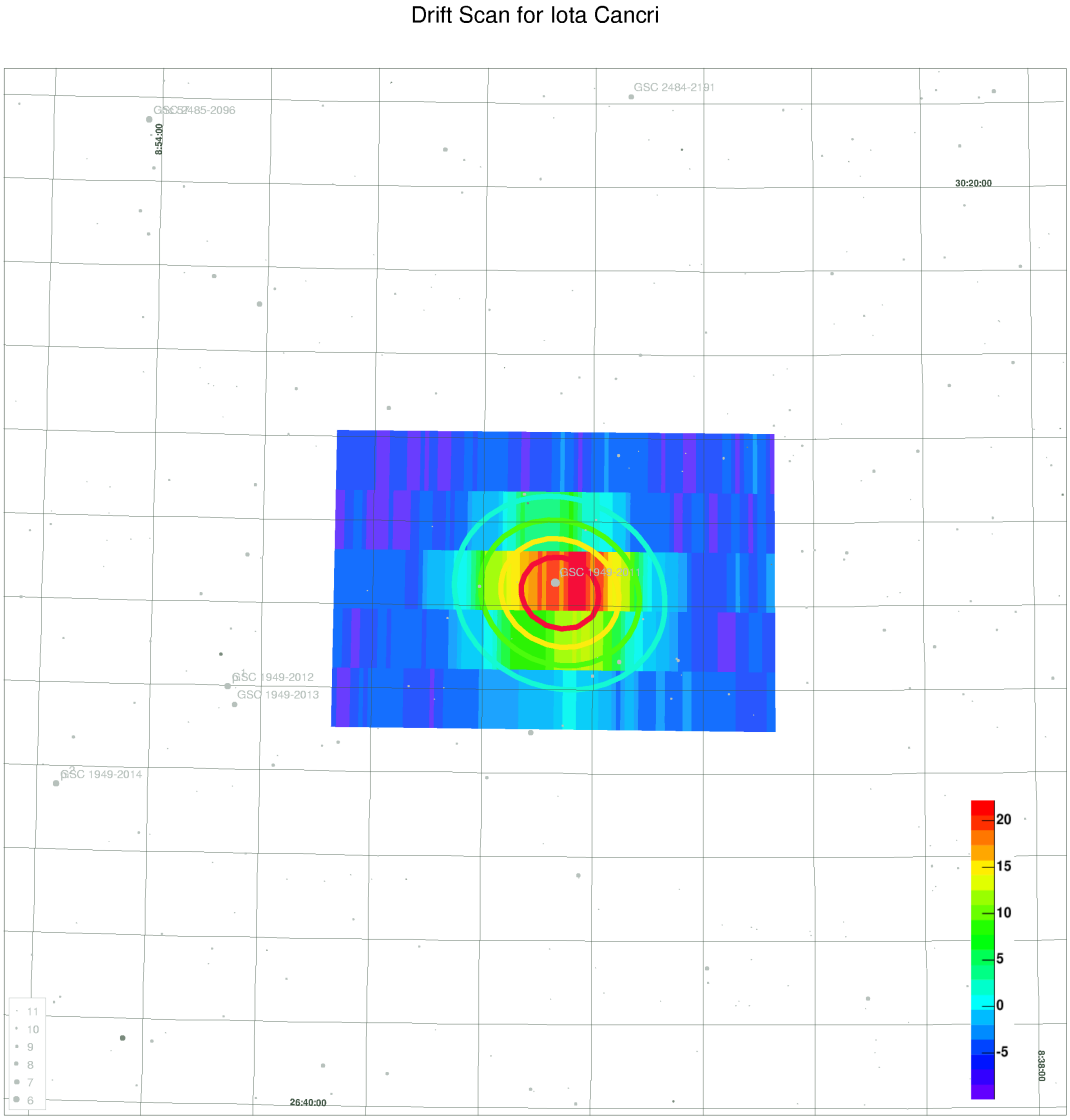

Figure 4.18: Typical PMT anode current distributions for a well-aligned heliostat during a drift scan on *Iota Cancri*, overlaid by a sky map of the surrounding area. The angular width of the area covered by the drift scan is 2° in Right Ascension (RA) and 1° in declination (δ), corresponding to five runs of 8 minute length, each offset by 0.25° in declination. The coloured rectangles indicate the PMT current for each (RA, δ) bin in units of μA, and the coloured ellipses are contour lines of a two-dimensional Gaussian fit to the current distribution. Taken from Bramel (2005).

Raster Scans

Before the drift scan method became the standard within STACEE, the heliostat biases were determined through procedures called *FBIAS* and *SBIAS*. In both techniques, the PMT anode currents were recorded while the heliostat biases were changed "on the fly" by varying amounts in both azimuth and elevation direction, following a predefined pattern. In the offline analysis, the resulting two-dimensional current distributions were fit with 2-dimensional Gaussians. The bias settings yielding the highest currents, given by the peaks of the Gaussians, were then found for each heliostat. As a light source, a bright celestial object was used, and tracked by the heliostats in parallel mode; the Moon in the case of FBIAS and a bright star for SBIAS. In both cases, each camera's cans were divided up into two interspersed halves in order to minimise optical cross-talk due to cans receiving light from images of heliostats associated with neighbouring cans.

While the accuracy of this method was comparable to that of the drift scans, the raster scan technique was given up in favour of the former because it was found to be extremely sensitive to the exact sky position of the target object.

Sunspots

The NSTTF personnel set the heliostat biases used for solar tracking during the day by projecting images of the Sun, close to solar noon, onto a target on the front (North) face of the Solar tower, and adjusting the biases until the images are well centered. A CCD camera is used to take images of the resulting light distributions, called *sunspots*, which are then analysed offline and used to confirm the spot position. At the same time, the lateral spread of the light distributions provides a qualitative measure of how well the heliostats are focussed. Sunspots were also used to align the individual heliostat facets before the laser alignment procedure was implemented (see Sec. 4.2.2), by adjusting their headings until the spot size reached a minimum. Furthermore, the CCD images of sunspots could be compared to the intensity distributions obtained with the STACEE optical (ray-trace) simulation.

Since this method was the least complicated and required no data from the DAQ (*e.g.* in the form of PMT currents), it was occasionally used by STACEE to set temporary biases whenever maintenance had been done on a heliostat and circumstances had not yet allowed the performance of drift scans.

The disadvantage of the sunspot method is a certain lack of accuracy due to the relatively large size of the images, and the fact that it did not take into account any effects caused by the STACEE optical system (secondary mirrors and PMT cans), such

as the translation of the inhomogeneous morphology of the sunspots due to the individual facets into light distributions on the focal planes of the cameras.

4.2.2 Facet alignment and focusing

As mentioned earlier, each heliostat's facets can be aligned and focussed individually. This was done manually while observing the light distributions due to laser light reflecting off individual facets onto a target at the north side of the tower about 10 metres above the 260-level secondary mirrors; see Fig. 4.19 for a sketch of the setup.

The laser, installed on a tripod on the tower roof, was equipped with a beam diffuser mounted in front of its aperture, allowing the beam width to be adjusted so that all of the target facet, but none of the neighbouring facets, was illuminated uniformly. For each heliostat facet, the heliostat was commanded to a heading at which the facet would reflect the light from the laser onto the target, if it were aligned in such a way with respect to the heliostat body frame that it would focus the light from a source at the so-called *tune-up heading* (azimuth 180°, elevation 70°) onto the appropriate secondary mirror. This procedure was performed at dusk or at night, so that the laser operator could see the light distribution due to the illuminated facet. The laser was aimed at the central facet first, and in case of a misalignment of the image greater than 5 cm with respect to the target, the heliostat was rebiased so that the central facet was at the

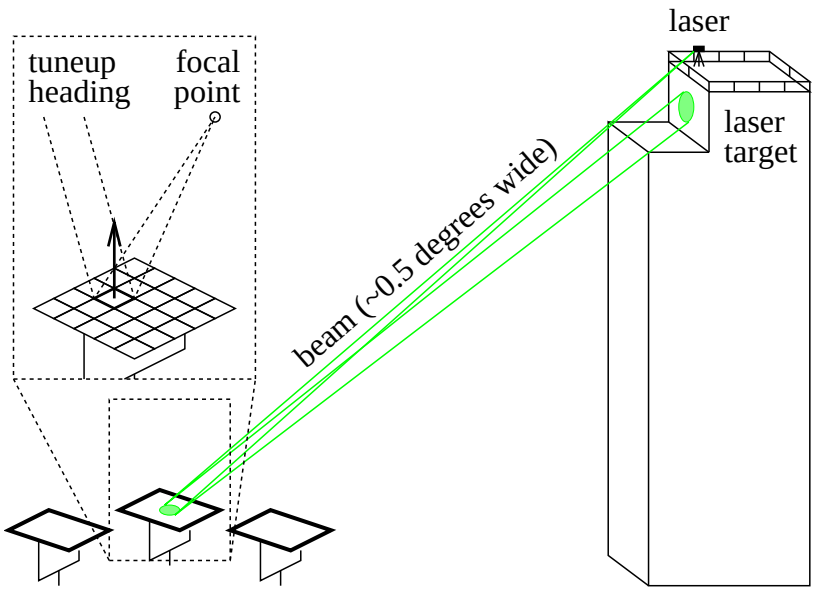

Figure 4.19: Sketch of the procedure used for heliostat facet alignment and focusing, by means of a laser positioned on top of the Solar tower. Taken from Scalzo (2004).

correct heading. (Note that there is a *caveat* attached to this part of the method, which is described in detail in Sec. 4.2.3.) After this, the remaining 24 facets were targeted by the laser and aligned with respect to the centre facet, via instructions given to a second operator on a hydraulic lift in the field beside the heliostat, who mechanically adjusted the facet heading until the image was centered on the target. Likewise, the foci of the facets were adjusted until the light distributions were smallest (or rather, until the largest fraction of the integrated intensity appeared to fall into a circular area the size of a secondary mirror). Due to imperfections in the facets, which include large departures from the parabolic shape as well as cracks, the images were usually filamented and often very irregular, exhibiting large spidery tails. This fact made the centering and focusing a somewhat difficult and subjective procedure, limiting the accuracy to which the facets could be aligned in terms of lateral distance from the target to between about 10 cm for the closest and 20 cm for the farthest heliostats.

4.2.3 Skewness correction

Results of raster scans performed as part of the heliostat alignment procedure (see Sec. 4.2.1 had suggested that for some heliostats, the pointing offset (in both magnitude and direction) was not constant, but depended on the absolute heading of the heliostat. The heliostats used at the NSTTF were not designed for tracking objects with sub-arcminute resolution. Because the Sun is an extended object and the original purpose of the facility rarely included using one heliostat alone, but rather the whole array ($\sim$220 heliostats), small pointing errors of individual heliostats did not pose a problem. Due to the relatively loose requirements on the accuracy of the heliostat pointing, imperfections in the drive mount as well as facet alignment can lead to a direction dependent pointing offset. The main causes are likely to be the following, in order of importance:

- global alignment of the facets with respect to the heliostat body frame
- deviation of the azimuth drive axis from vertical
- different lengths of the two fork arm tines

The latter two, if significant, were difficult to correct for. For astronomical telescopes, where exact pointing is much more important than for atmospheric Cherenkov telescopes, pointing models containing a set of *misalignment parameters* are used to calculate corrected pointing directions and thereby account for such imperfections. For an array of 64 heliostats, this would not just be impracticable but also unnecessary.

A global facet misalignment, on the other hand, can potentially introduce a direction dependent pointing error of up to several bits, which would be a problem even for imaging relatively large regions such as Cherenkov light showers. At STACEE, such a situation could occur because in the facet alignment procedure described in Sec. 4.2.2, it was assumed that the optical axis (*i.e.* the normal) of the *central* facet of each heliostat was perfectly perpendicular to the elevation axis. In reality however, that facet is mounted to the heliostat frame in the exact same way as the other 24 facets, *i.e.* with adjustable heading in both azimuth and elevation direction. Although during normal use the central facet heading was never changed, a certain degree of misalignment could be expected with time due to thermal expansion and contraction, vibrations of the mount during normal heliostat operations, *etc.*

In order to measure and correct a misalignment of the central facet, the setup for the laser-aided facet alignment procedure as described in the previous section was used. There were two azimuthal headings of a heliostat which result in the image being on target: the nominal heading, and an "inverted" heading $\approx 180^\circ$ offset from the former, with the heliostat driven past the 90° point in elevation (*i.e.* flipped). Comparing the azimuthal headings for both cases yielded the central facet misalignment: in the ideal case, the difference should be exactly 180°, and any deviation from this value had to be due to a non-orthogonality between the facet normal and the elevation axis. The azimuthal facet heading was then corrected by half the measured offset value in the correct direction, which brought it back to the square position with respect to the heliostat body frame.

The central facet misalignment was measured on all 64 STACEE heliostats in the fall of 2002, and corrected on any heliostats with a misalignment of more than 2 bits in the time between September 2002 and January 2003. A misalignment of less than 2 bits was considered negligible.

4.2.4 Camera alignment

The positions of individual cans on the camera plates were determined by tracking the full Moon with one heliostat at a time. A small white target placard was attached to the upper end of the corresponding can, and a CCD image of the light distribution was taken. By analyzing the images offline, the exact can positions which maximise the integrated light intensity falling on the apertures of the cans were found. A second set of images taken after readjustment of the can positions was used to verify that the light collection had improved to within 5% of its optimal value.

As for heliostat biasing by means of raster scans, camera alignment was performed when the Moon was close to transit, and preferably in the winter months due to its higher elevation. This is because off-transit positions increased the amount of spherical aberration, leading to long coma tails such as the ones seen in Fig. 4.16, making it harder to determine the optimum position of the cans. It is interesting to note that for some heliostats, the tails were long enough to overlap with the apertures of neighbouring cans. While the farthest heliostats, which were the most *off-axis*, experienced a certain degree of image distortion even for optimal headings, this was not a problem for normal astronomical observations: the arrival times of the Cherenkov wave front generally differ by tens of nanoseconds even for neighbouring heliostats, causing such *cross-talk* photons to arrive outside the trigger coincidence window.

4.3 Heliostat and weather monitoring

During normal data taking, both heliostat status and weather data were continuously recorded. Furthermore, the sky clarity in the field-of-view was measured via photometry on designated stars close to the source position. The data acquisition for these subsystems was separate from that of the main STACEE data, which will be described in Sec. 4.4.6, due to difficulties caused by the fact that each of the three subsystems ran on a different computer platform, and, to a lesser extent, because they were located in different buildings: the heliostat control and weather monitoring systems were housed in the heliostat control tower to the north of the heliostat field, while the main DAQ was located in the Solar tower along with the PMT cameras. This section provides a brief description of the peripheral systems which were located in the heliostat control tower.

4.3.1 Heliostat status monitoring

The heliostat control system, also called *Master Control System* (MCS), ran on a Hewlett-Packard 2000 mainframe computer from the early 1970's. A graphical display showed the heliostat status including tracking or slewing mode, assigned target, current pointing coordinates (in bits) and any tracking or communication errors for all heliostats. This information was logged to disk every 20 seconds and, at the end of observing nights, transferred to a DOS computer via serial link and subsequently merged with the rest of the STACEE data. While the heliostat status information in the data stream is essential for offline analysis, the MCS online monitoring system allowed the heliostat operator to make conscious decisions about deactivating malfunctioning heliostats or attempting to rectify the problem *in situ*. If a heliostat suffered from consistent tracking errors (course

track, fine track or drive oscillations), it was generally stowed and deactivated for the remainder of the night's observations.

4.3.2 Weather and atmospheric monitoring

STACEE's weather and atmospheric monitor system comprised two components: a weather station and a telescope-based stellar photometer. Both systems were controlled via RS-232 serial links from a Windows PC using a graphical data acquisition interface written in *LabView*.

The weather station continuously wrote local temperature, humidity, wind direction and wind speed to disk. Acoustic warnings informed the operator if the wind speed was too high for the heliostats to be operated safely (> 35 mph) or if there was a risk of lightning strikes. A quantity called *frost index*, defined as the difference between temperature and dew point, was also calculated, and used both during observations and in the offline analysis to assess the probability that frost or dew had formed on the heliostat facets.

The purpose of the STACEE photometer (for very detailed descriptions, see Scalzo, 2000b, 2002) was to get an independent measure of the sky clarity (*i.e.* atmospheric extinction) by measuring the difference between the apparent brightness of suitable calibration stars and that of the surrounding background sky field. For the star tracking and light collection, an 8" *Meade LX200* Schmidt-Cassegrain telescope on a *Losmandy G-11* German equatorial mount was used, connected to a Losmandy Gemini positioning system. Attached to the telescope barrel was an *Electrim EDL-1000L* CCD camera with a telephoto lens, which was used by the autoguider to periodically correct the pointing of the telescope.

Until the fall of 2004 – and thus for all the observations relevant to this thesis – the photometry was performed by means of a *Hamamatsu H3460-53* photomultiplier mounted to the exit aperture of the telescope. The PMT had a narrow field of view of 12 arcminutes half-angle, making its response very sensitive to slight pointing inaccuracies. In the beginning of the 2004/2005 observing season it was therefore replaced by an *SBIG* CCD camera. The chosen calibration stars were non-variable stars of magnitude 2 to 4, located within about three degrees of the potential sources, and were monitored with the photometer while the corresponding sources were observed with the gamma-ray detector. Due to various difficulties with the operation, such as insufficient pointing accuracy in the case of the PMT, as well as with the interpretation of the photometry data, the latter were never actually used to for applying quality cuts to the Cherenkov

data. The atmospheric monitoring system also included two infrared cloud detectors, one with a 90° field of view pointing at zenith, and one with a 3° field of view mounted alongside the telescope, *i.e.* viewing the same region of the sky as the heliostats. The voltage outputs of the cloud detectors were written to disk in text format. Unfortunately the cloud detectors were found to be even less indicative of the actual sky conditions than the photometer, and are therefore not used in the data analysis.

4.4 Electronics

The detection of the short-duration (few nanosecond) Cherenkov light flashes from EAS requires detection techniques very different from those of conventional optical or radio telescopes. In fact, the electronics used to read out atmospheric Cherenkov telescopes are very similar to those used by particle physics detectors. Apart from the quick response times necessary to detect and reconstruct an EAS, discrimination capabilities against the overwhelming background of night sky photons are required. At STACEE, this goal was accomplished by using fast photomultiplier tubes (PMTs) combined with a multiplicity trigger and *Flash-ADCs* (FADCs) with nanosecond resolution. In the following we describe the STACEE electronics setup, and examine its main components in some detail.

4.4.1 Photomultiplier tubes

The requirements on photomultiplier tubes used in Air Cherenkov telescopes – and for STACEE in particular – are quite strict. First and foremost, they should have fast response times and narrow output pulses so that the thin (around 5 ns $\cdot\, c$) Cherenkov wavefront can be detected over the high background of random fluctuations caused by night-sky photons.[2] Furthermore, a high quantum efficiency at blue and near-ultraviolet wavelengths (*i.e.* the part of the spectrum where the bulk of the Cherenkov light in EAS is emitted) is desirable. Another relevant property is the internal jitter in propagation time of a pulse through the dynode stages of a PMT, the RMS of which is called the *transit time spread*. For Cherenkov telescopes the latter should be as small as possible, as timing jitter will degrade the time resolution of the experiment by spreading out the time profile of a Cherenkov shower and, just as in the case of wider output pulses, make it harder to discriminate it against the night sky background. Finally, the tubes need to

[2]This was done, both online in the trigger and offline in the reconstruction, by applying a narrow coincidence gate on all channels and requiring a multiplicity condition to be satisfied. See Sec. 4.4.4 for details.

be able to operate in an environment with high night sky background flux. At STACEE, due to the large collection areas and fields-of-view of the individual heliostats, light from stars alone generally gave rise to PMT anode currents in the micro-Ampere range. Light from man-made sources, scattered back to the ground by water droplets and aerosols in the atmosphere, *i.e.* light pollution, added to the natural background. Since the NSTTF is located only 10 km from the edge of the city of Albuquerque, scattered artificial light was by far the dominant source of background. The background PMT currents due to both starlight and artificial light routinely exceeded 10 μA and, even in conditions with excellent sky clarity, were generally in the range between 20 and 50 μA, which means the primary photoelectron rates were in excess of 1.5 GHz.

The STACEE experiment used PMTs of the type *Photonis XP2282B* (Photonis PMT catalogue, 2004; www-Photonis, 2005). These phototubes are 51 mm in diameter, equipped with a bialkali (Sb-K-Cs) photocathode and have eight focussed dynode stages, making them sensitive to light with wavelengths between 290 and 650 nm, with a peak quantum efficiency of 28% at a photon wavelength of about 380 nm (see Figs. 4.17 and 4.20). The entrance window is made of borosilicate glass, which has good transmission to UV light with a cut-off of about 270 nm. The PMT is reasonably fast, having a rise time of 1.5 ns and a single photoelectron pulse width of 2.2 ns (FWHM). In part because of the small size of the PMT, the transit time spread is only 0.4 to 0.5 ns. The chosen *VD182K/C* tube base housed a transistorised voltage divider, which improved linearity at high incident photon fluxes by keeping the dynode voltages constant over a wide range of anode currents. With the base, the XP2282B was "certified" to operate at currents up to 180 μA, which made it suitable for operation in STACEE's environment.

Typical PMT supply voltages were around -1600 V, yielding tube gains of around $(1.0 - 1.3) \cdot 10^5$. These values were on the lower end of the range recommended by the

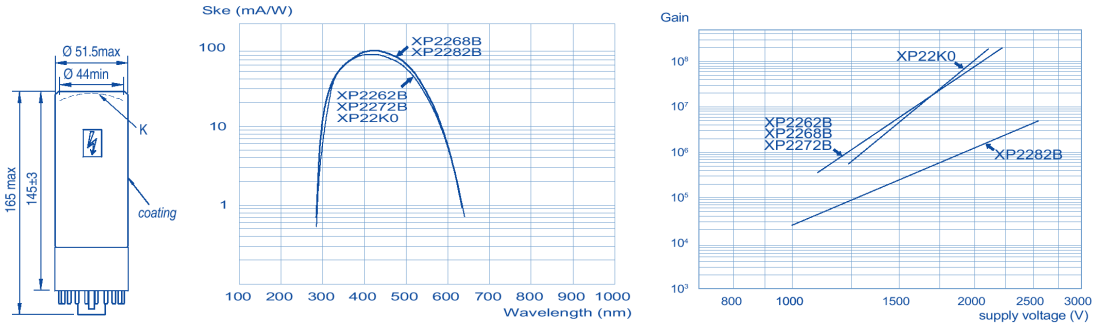

Figure 4.20: Dimensions (in mm), spectral response and typical gain of the Photonis XP2282B photomultiplier. From Photonis PMT catalogue (2004).

manufacturer, and had been chosen to minimise the wear on the PMT photocathode resulting from extensive operation at high currents. The high voltage was supplied to the PMTs by two *LeCroy 4032A* high voltage power supplies, controlled remotely from the control room on the 200-level by a *LeCroy 2132* CAMAC interface.

4.4.2 Upstream electronics

Fig. 4.21 outlines the electronics setup for the STACEE experiment, summarizing the signal processing and trigger electronics downstream of the PMTs. Detailed block diagrams including all modules can be found in Figs. 4.22 (all electronics upstream of the trigger) and 4.23 (trigger and veto circuit).

The analog signal from each PMT propagated through 35' of shielded RG58 signal cable to an electronics rack on the camera platform. On the 260-level, this rack was located in a metal enclosure for RF-shielding purposes. There, it passed through an RC high-pass filter box designed to block out the DC contribution. These boxes also contained a voltage divider, by means of which part of the incoming signal was branched off and sent down to the control room via 150' long ribbon cables, where it was fed into *Joerger* scanning ADCs for DC current monitoring. The AC-coupled signal was sent through two successive stages of *Phillips Scientific 776* amplifiers with a nominal bandwidth of 275 MHz, each of which amplified the signal by about a factor of 10. From a patch panel, the amplified signal was sent down to the 200-level electronics in the control room through 150' of RG213 cable, where it connected to a linear fan-out. Two (identical) outputs of the fan-out were passed to the signal input of an *Acquiris* Flash-ADC (see Sec. 4.4.5 for details) and a 16-channel constant-threshold discriminator, respectively. The discriminator thresholds were set remotely by the DAQ via a CAMAC interface. The voltage to which they were set, which corresponded to roughly six photoelectrons, was identical throughout all 64 channels but varied somewhat from run to run or night to night, depending on the location of the breakpoint in the rate-versus-threshold curve (see Sec. 4.5.4). The length of the discriminator output pulses was set to the lowest possible value (around 6 ns), in order to minimise the discriminator deadtime.

The logic NIM pulse output of the discriminator was sent through a ribbon cable to the corresponding input of one of the cluster boards of the delay and trigger system, which will be described in detail in Sec. 4.4.4; another output was connected to a *PS 7132* scaler used to monitor the PMT discriminator rates. The inputs of two 32-channel *LeCroy 3377* common-stop TDCs were connected to the end of the same ribbon cable that serviced the inputs of the trigger boards. TDCs as well as FADCs were read out by

the DAQ whenever an event trigger occurred, which could be of three different types: real, fake or calibration. The trigger circuit will be described in the following section.

4.4.3 Trigger circuitry

A block diagram of the trigger circuit is shown in Fig. 4.23. An event trigger could be generated in three different ways. The digital trigger system (commonly called *MADDOG*; see Sec. 4.4.4) continuously tested the incoming NIM pulses from the PMT discriminators for a two-level coincidence condition, and was designed to trigger – after application of the appropriate delays – on Cherenkov as well as laser events. The triggers generated by MADDOG are commonly referred to as *real* triggers. A *fake* trigger source (a *PS 794* gate/delay generator), normally set to 0.5 Hz, periodically "faked" a trigger signal, prompting a module readout which was used offline for pedestal determination. Finally, a trigger could come directly from the the laser calibration system, which we shall describe in Sec. 4.5.1. These triggers are called *calibration* triggers.

Each type of trigger signal was fanned-out by a separate *Phillips 755* majority logic unit, which got inhibited through a feedback loop via a veto input whenever a trigger had occurred, so that additional triggers did not interfere with the event readout. Two outputs of each of these units were connected to scalers, used as event counters. The third output led to a common trigger fan-in/out, which, upon receiving any of the three trigger signals, distributed NIM pulses to several modules. Via a set of fan-outs, one copy was sent to the external trigger inputs of the FADCs, prompting the readout of the FADCs. Another copy was sent to the common stop inputs of the TDCs via a NIM-ECL converter module. A GPS clock, accurate to $1\,\mu$, also received a copy of the pulse after it has passed through a NIM-TTL converter, triggering a GPS clock readout. A final copy was sent to a flip-flop (a *Phillips 794* gate generator), which generated the event veto signals for the three *PS 755* modules and output a veto for a deadtime counter. The latter consisted of a 10 kHz clock, whose output was split up into a vetoed and an unvetoed copy, which were then counted by two scalers. When the module readouts (MADDOG, FADCs, TDCs, GPS clock) had completed, the trigger condition was cleared by the DAQ, which used an output register to reset the flip-flop. At that point the system was ready to accept a new trigger.

4.4.4 Delay and Trigger System

In principle, the first stage of the STACEE trigger were the discriminators, which produced output pulses whenever the analog input voltages exceeded a set threshold, which

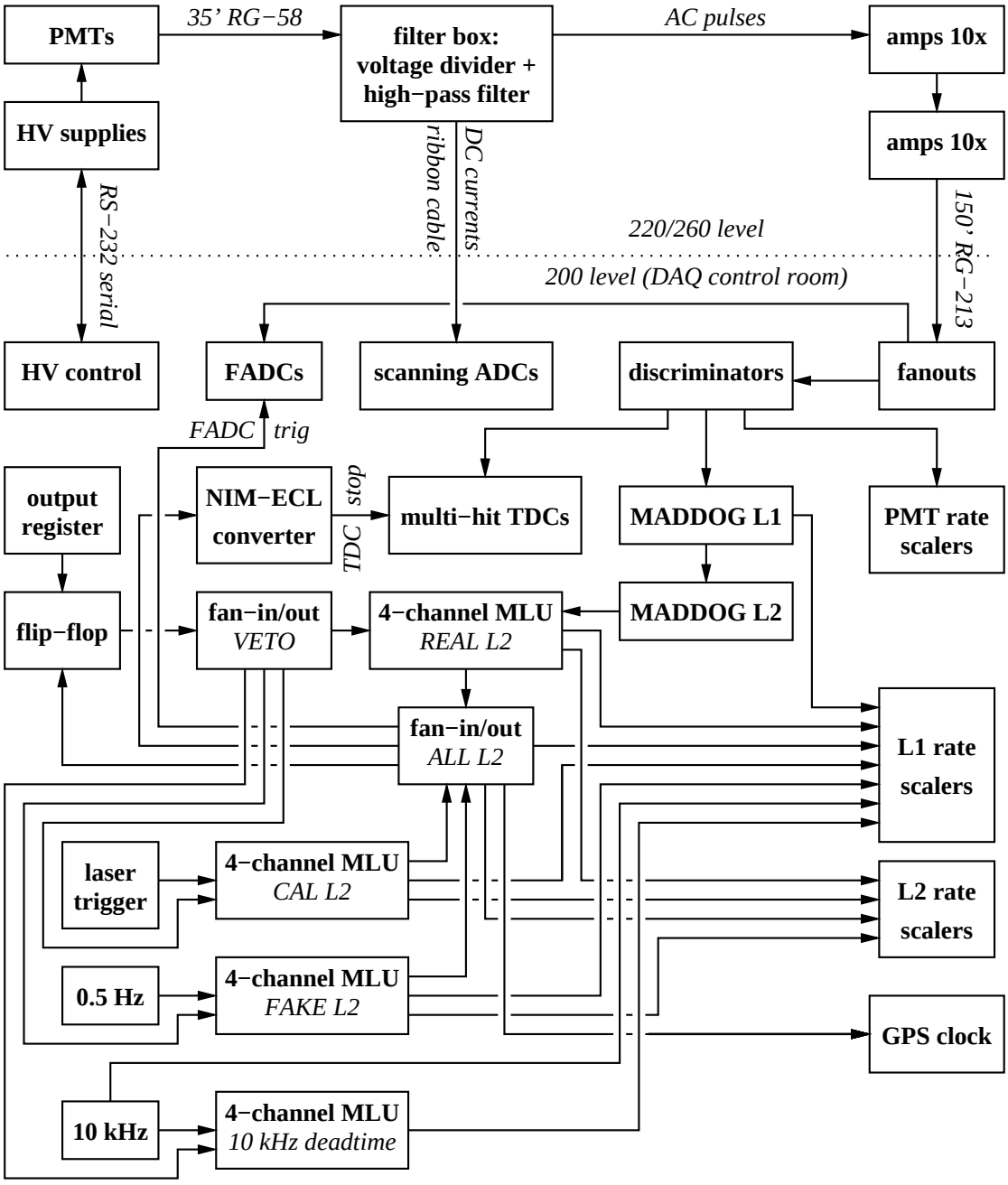

Figure 4.21: Schematic diagram of the complete STACEE electronics setup. Incoming photons are converted by the PMTs to analog electronic signals, which are then AC-coupled and amplified. Fan-outs produce two copies each of the amplified signals, which are sent to the FADC system and a discriminator, respectively. The logical pulses from the discriminators are fed into the delay and trigger system (*MADDOG*), which applies the appropriate delays and sets the coincidence requirement.

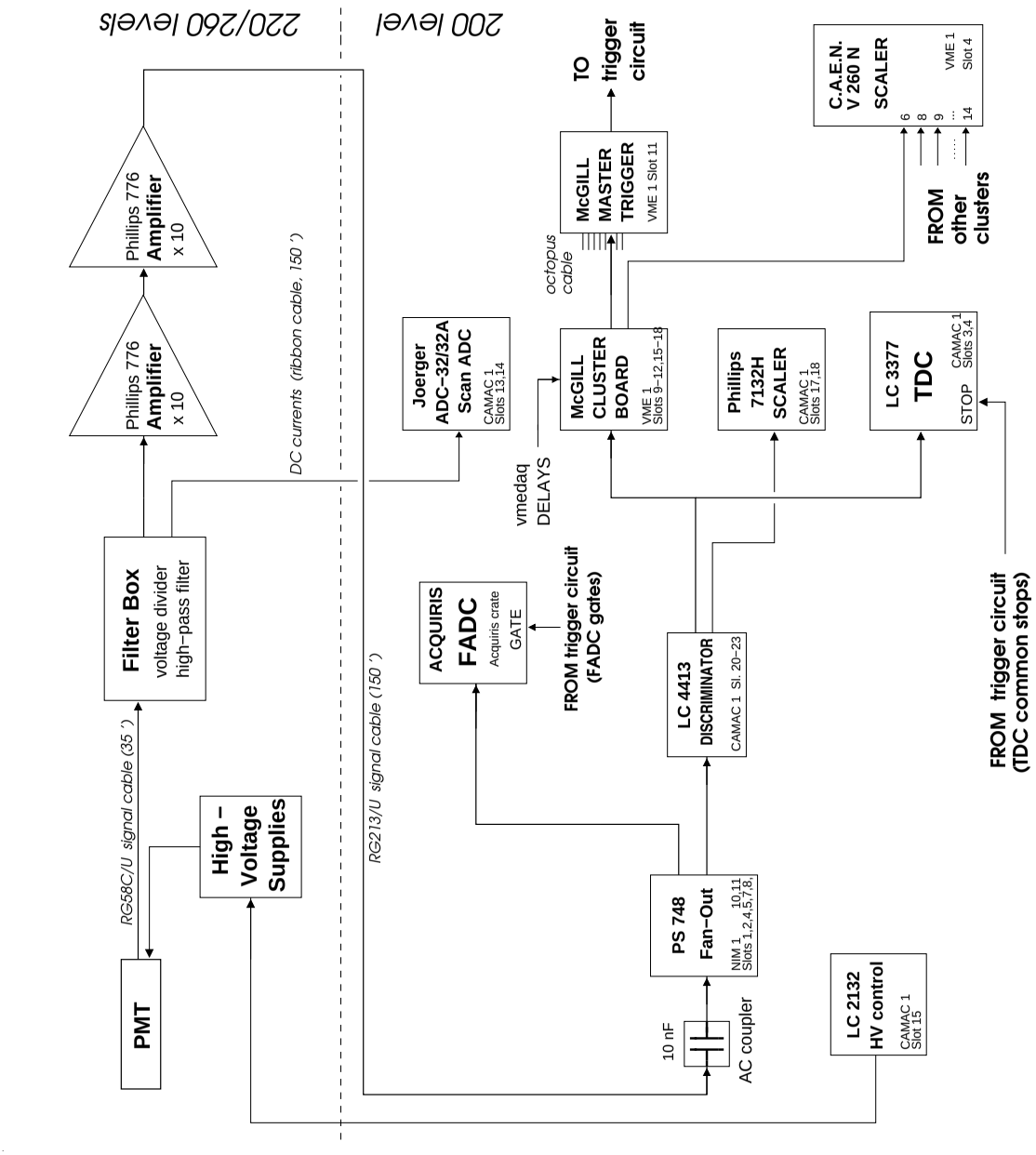

Figure 4.22: Detailed block diagram of the upstream part of the STACEE electronics. The trigger circuit is depicted in Fig. 4.23.

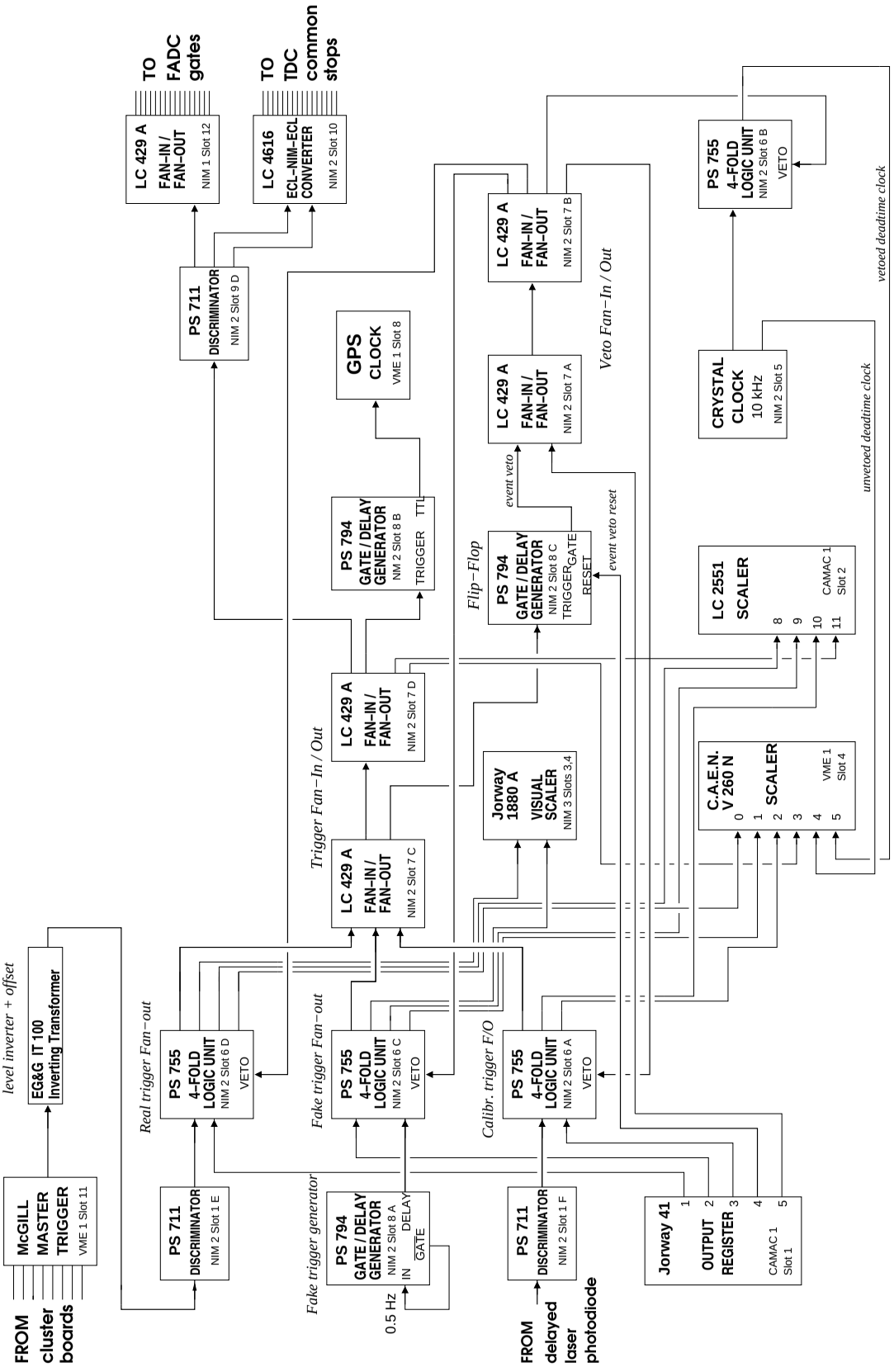

Figure 4.23: Detailed block diagram of the downstream part of the STACEE electronics setup, including trigger circuit.

is usually kept constant throughout all 64 channels and corresponded to roughly 5 to 6 photoelectrons per channel. The MADDOG delay and trigger system was responsible for applying appropriate delays to the discriminated signals from all 64 STACEE channels, and for triggering data readout whenever a given two-level coincidence condition was satisfied.

As detailed in Chapter 3, Cherenkov light showers from gamma rays attain a roughly spherical shape at ground level, exhibiting a narrow timing profile with a width of only about 5 ns. The detection of these brief Cherenkov flashes, amongst the overwhelming background of night-sky background fluctuations as well as Cherenkov showers due to hadronic primaries incident from random directions, requires a trigger system with nanosecond resolution which is sensitive to the expected shapes of Cherenkov wave fronts produced by air showers from gamma rays coming from the observed source. The latter was achieved by applying a time "mask" to the signals from the individual PMTs by subjecting them to channel- and time-dependent delays. As the Cherenkov wavefront was sampled at 64 different positions on the ground, the times of arrival at the individual heliostats differed by tens of nanoseconds, depending on the exact shape of the wavefront and – in the case of gamma rays – the direction to the expected gamma-ray source. This direction dependence, which is unique to wavefront sampling Cherenkov telescopes and arrays of imaging ACTs, required the delays to be updated dynamically during observations as the source tracked across the sky. The relative signal delays were chosen such that a spherical wavefront emanating from the canting point, which is where the expected shower maximum was located, caused the signals from all 64 channels to arrive at the trigger simultaneously. Due to the relatively large width of the trigger coincidence window (around 24 ns), this assumption was reasonable for identifying gamma rays; *e.g.* it allowed showers with more conical wave fronts to pass the trigger, as well as showers whose maximum occurred at different – albeit reasonably close – heights than the nominal canting height.[3]

The applied signal delays had three components. Apart from the direction- (and therefore time-) dependent *source-to-heliostat* delays described above, the signals from each channel were subjected to a constant delay to account for the differences in path length between individual heliostats and their corresponding secondary mirrors. These delays are called *time-of-flight* (TOF) delays. Finally, there were channel-to-channel corrections, called *trims*, for slight differences that existed in the propagation times through the electronics, which were in the order of one to two nanoseconds. While the

[3]This is important since, as we have seen in Chapter 3, there is a strong correlation between the mean height of shower maximum and the energy of the gamma-ray primary.

former two contributions were calculated by the DAQ program, the trims were read in from a file which was kept up-to-date via periodic measurements using the laser calibration system.

Both the delays and a two-level trigger coincidence condition were imposed by a custom-made digital delay and trigger system called *McGill Asynchronous Digital Delays for Observation of Gammas* (MADDOG), which was based on *Altera FLEX 10KE* embedded programmable logic devices (Martin & Ragan, 2000; Ragan, 2003). It consisted of 10 VME logic boards; a clock board, eight 8-channel cluster trigger boards and a global coincidence trigger board. The clock pulses from the 32 MHz TTL clock oscillator were distributed via the VME bus to the other nine boards, which used it to synchronise locally generated 64 MHz and 128 MHz clocks. In the cluster boards, the channel-specific delays were applied, and subsequently the delayed signals of the eight channels within each cluster board were tested for coincidences within a series of coincidence gates of predefined width. If the number of channels with coincident "hits" exceeded the intra-cluster *local trigger condition* (LTC), the cluster board sent out a logic signal to the master trigger board, which acted as a simple multiplicity logic device by imposing the *global* trigger condition (GTC) on the inputs from all eight cluster boards.

In the following, the way in which MADDOG performed the delaying and triggering will be described in more detail. The discriminator pulses belonging to the eight channels of each heliostat subcluster as defined in Table 4.1 (see also Fig. 4.6) were fed into the same cluster board, where they were digitised at 64 MHz with a resolution of 1 ns. The latter was achieved by *tapping off* the incoming signals onto a set of eight parallel delay lines for each 128 MHz clock cycle, producing a bit pattern that was subsequently encoded. These so-called *vernier codes* consisted of four bits: three bits for encoding the number of the delay tap which coincided with the rising edge of the incoming discriminator pulse, and, since *zero* was a valid time code corresponding to the first delay tap, one extra bit was needed as a flag to indicate a "hit" (*i.e.* the arrival time of a discriminator edge within the given interval). This means there could be at most one hit per 7.8 ns interval. The synchronisation with the 64 MHz clock was done by splitting up the 15.6 ns clock pulse into two separate paths with a relative delay of ~ 8 ns, and resynchronising the two 4-bit vernier codes by writing them into a register as an 8-bit code at the following 64 MHz clock edge.[4]

[4] At the time the boards were designed (1999), the availability of field-programmable gate arrays that could be clocked at frequencies as high as 128 MHz was quite limited. This is the reason why a 64 MHz clock was chosen instead.

Cluster	Channels	Tower Level	Camera
0	00 – 07	260	East
1	08 – 15	260	East
2	16 – 23	260	West
3	24 – 31	260	West
4	32 – 39	260	North
5	40 – 47	260	North
6	48 – 55	220	Southeast
7	56 – 63	220	Southwest

Table 4.1: Correspondences between STACEE channels, trigger clusters, PMT cameras and secondary mirrors.

It should be noted that although the delay taps had a nominal delay of 1 ns, the error due to the manufacturing process was quite large (0.5 ns), and therefore the occupancies of the individual delay taps were not uniform but showed a hardware-dependent pattern. Furthermore, since a 128 MHz clock was used, the total time available before the next clock edge was actually 7.8 ns rather than 8 ns, and therefore the last delay tap had in average a lower occupancy than the previous 7 taps. A typical example of the resulting vernier bit patterns is shown in Fig. 4.24. Due to the small magnitude of this effect in comparison with the timing resolution of the electronics, it can safely be neglected.

The actual delaying of the vernier doublets, each consisting of eight bits, was done by first appending a seven bit reference time, and then adding the 11 bits of programmed

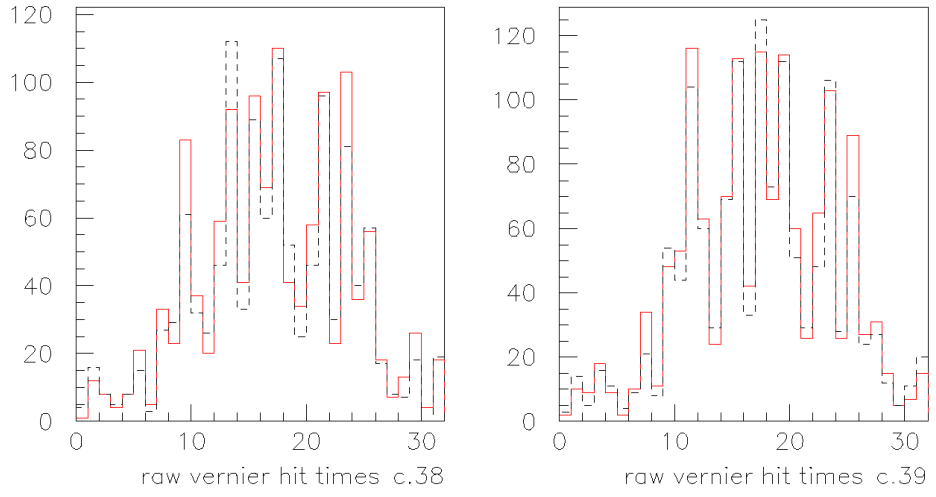

Figure 4.24: Delay tap occupancies for two zenith runs (runs with constant delays) and two representative channels. The pattern is a result of the unevenly spaced delay taps.

delay stored in the delay register of the channel in question. The resulting 11 bits of delayed time stamps were deposited in a *first-in first-out* (FIFO) memory which was 2048 ns deep. The upper seven bits of the first available (*i.e.* earliest) vernier doublet were continuously compared with the reference counter value. Whenever they matched, the doublet was inserted into a coincidence pipeline, along with the following doublet. This allowed there to be four vernier codes, or 31.25 ns worth of hit information, to be present at the coincidence pipeline for each of the eight channels of the cluster. The delays were recalculated by the DAQ and updated every few seconds as the source tracked across the sky.

At this point, the cluster trigger algorithm inspected the eight delayed time stamps for coincidences. The MADDOG boards were able to use two different methods to look for local (intra-cluster) coincidences, selectable with a software switch via the DAQ. In *wide mode*, the eight 32-bit patterns in the coincidence pipelines were tested for coincident hits within the first 24 ns *once* for each 64 MHz clock cycle, *i.e.* once every 16 ns. Due to the finite step width, full efficiency was reached only for time differences of 8 ns or less, resulting in a trapezoidal two-channel coincidence curve as shown in Fig. 4.25, which starts to rise at a time difference of 24 ns but reaches 100% only at 8 ns. On average, the wide coincidence mode was therefore equivalent to a coincidence window with a width of 16 ns. A cluster trigger was produced if the number of channels with at least one hit in this window equalled or was larger than the set local trigger condition (LTC).

In *narrow mode*, a series of coincidence windows with a width hardwired into the firmware (PROMs) was asserted in 1-ns steps, *i.e.* 16 times for each leading edge of the 64 MHz clock, each window starting one nanosecond later than the previous one. For most of STACEE operations, the width of the narrow mode coincidence gate was 12 ns, and the resulting two-channel coincidence curve looked like the one at the bottom of Fig. 4.25.

Originally, the wide coincidence mode was meant to be used for testing purposes only. However, due to worries that the narrow mode might not be functioning properly, it became the default mode for Cherenkov observations over several years. All data used for this analysis were taken in wide mode.

For a common LTC of 5, the local trigger rates, also known as L1-trigger or cluster rates, were in the order of between a few kHz and about 1 MHz. Obeying the laws of combinatorics, they depended sensitively on the eight input discriminator rates, which were between 1 and 20 MHz. Whenever a local trigger occurred, the cluster board output a differential ECL signal, coincident with the following 64 MHz clock edge, which was

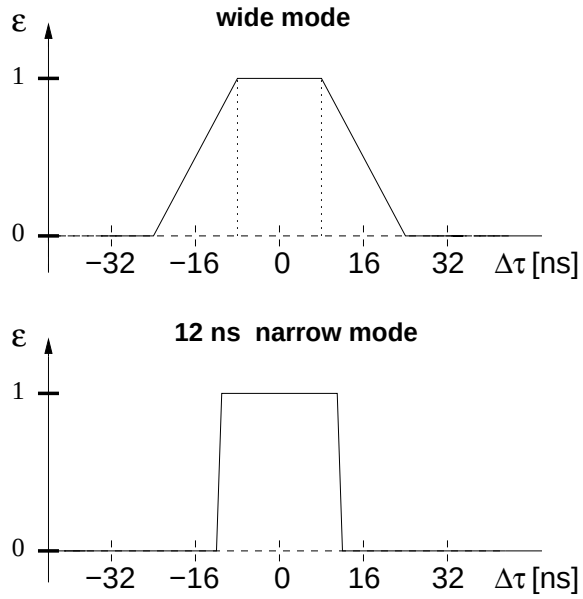

Figure 4.25: Two-channel coincidence curve for wide and narrow (12 ns gate) trigger modes as imposed by the MADDOG trigger system. ε is the efficiency.

routed to the master board through front-panel connectors. For each clock cycle, the master board examined the L1 trigger signals from all eight clusters, and generated a global (L2) trigger whenever the arithmetic sum of the L1 triggers equaled or exceeded the set global trigger condition (GTC). When this happened, halt signals in the form of ECL pulses were sent back to the cluster boards, preventing each board from producing further L1 triggers and allowing the 32 nanoseconds of vernier information to be read out. Furthermore, the master board sent a NIM pulse to the trigger circuit (see Fig. 4.23), which provided the FADCs and TDCs with stop signals and indicated to the DAQ that the event should be read out. Typical L2 rates, *i.e.* cosmic-ray trigger rates, were in the range of 5 to 7 Hz, depending on source elevation, sky clarity and the exact choice of discriminator thresholds.

The selection of the two-level trigger conditions, *i.e.* LTC and GTC, was critical to the experiment, as they not only affected the energy threshold, but also the ratio of triggers due to to gamma-ray initiated air showers to those from hadronic air showers. Simulation studies were been done in order to determine proper LTC and GTC settings; the optimum combination of these numbers was the one that maximised the number of low-energy gamma rays passing the trigger conditions, while at the same time maximizing the *quality factor* Q defined as

$$Q = \frac{R_\gamma}{\sqrt{R_{had}}} \tag{4.1}$$

where R_γ and R_{had} are the trigger rates due to air showers from gamma rays and hadrons, respectively. The former was necessary to achieve a low energy threshold, while the latter was directly proportional to the detector's gamma-ray sensitivity. Although the optimum (LTC, GTC) combination appeared to vary somewhat depending on the exact values of certain detector quantities such as the PMT gains, or whether the mono- or paracanted heliostat scheme was used, simulations generally seemed to favour $LTC = 5$ and $GTC =$ 6 or 7. In practicality, however, other factors came to play that restricted the range of useful trigger multiplicity values. The good acceptance to low-energy gamma-ray showers of the stricter trigger conditions was due to low PMT discriminator thresholds, which in turn were possible because of low breakpoints (for a definition of this term, see Sec. 4.5.4). The low thresholds on the other hand gave rise to very high PMT discriminator rates and thus increased deadtime. Furthermore, the resulting high L2 trigger rates were suspected to cause stability problems with the FADCs due to the increased readout frequency. Finally, high values of LTC left the door open to a serious degradation of the detector sensitivity in the case of malfunctioning heliostats. The trigger multiplicities were thus left constant at conservative values of LTC = 5/8 and GTC = 5/8 for all STACEE observing, providing a reasonable compromise between low energy threshold and detector stability.

4.4.5 Waveform digitisers

Using only the timing information from the trigger system and the TDCs for the selection of gamma-ray events provided STACEE with a raw efficiency that was sufficient only for detecting the brightest gamma-ray sources, such as the Crab Nebula, by comparing the number of triggered events in the on-source to those in the off-source runs. Improving the raw sensitivity could only be done by rejecting Cherenkov showers of hadronic origin or other background beyond the level provided by the hardware trigger. The only way of accomplishing this was by looking at the triggered events in more detail, and studying properties such as the integrated charges or the rise and decay times of Cherenkov pulse candidates.

For this purpose, a set of 64 waveform digitisers, or *Flash-ADCs* (FADCs), was used. For all triggered events, the FADCs digitised the signals of all channels and stored them on disk, which were then examined in the offline analysis. As shown in Fig. 4.22, the analog input signals for the FADCs were tapped off directly from the fan-outs. For each triggered event and each of the 64 channels, an interval of 192 ns (out of a total of 2048 ns) of the input signal was digitised and recorded, centered around the expected

position of a pulse due to a Cherenkov shower wavefront from the direction of the observed astrophysical source (*i.e.* corrected for the delays). A section of 400 ns *before* the recorded portion of the input signal was also digitised and used to calculate a variance and a pedestal (baseline value) for each event and each channel. These values were stored with the trace data, and allowed inferred quantities such as integrated charges and pulse heights to be corrected for fluctuations in the background levels on an event-by-event basis.

The FADCs used for STACEE were *Acquiris DC270* with a sampling rate of 1 Giga-sample per second and a resolution of 9 bits, providing sufficient time resolution for characterising Cherenkov pulses. For normal observations a full-scale range of 1 V was chosen, compromising between dynamic range and voltage resolution. Several offline algorithms for reconstructing pulses from large showers that saturate the FADCs exist. A sample FADC trace of a typical Cherenkov pulse is shown in Fig. 4.26.

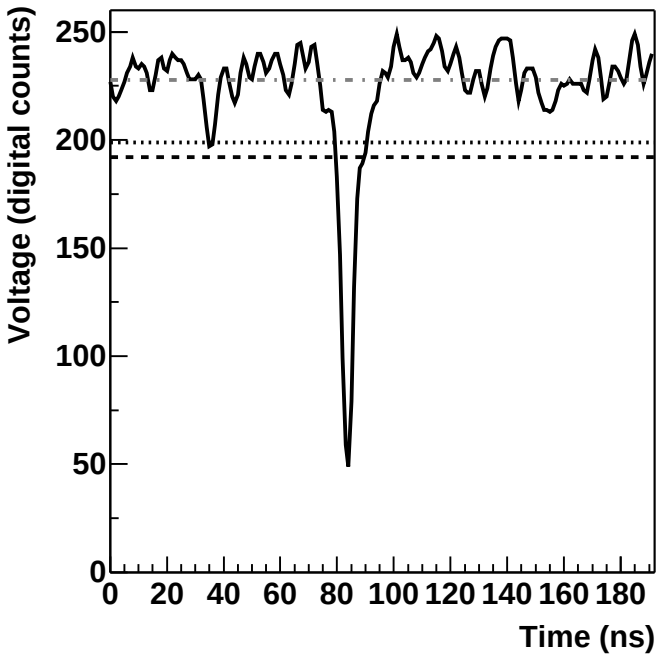

Figure 4.26: Sample FADC trace from a typical Cherenkov air shower event. The dashed line corresponds to the discriminator threshold and the dot-dashed line to the FADC baseline. The dotted line is the level of night-sky background fluctuations at three standard deviations above FADC baseline.

The 64 STACEE channels were distributed over four 16-channel Acquiris FADC crates. The DAQ program for the FADCs ran under Linux and was written by the

STACEE collaboration, including a Real Time Linux driver for fast writing of the trace data to disk (Schuette, 2000). During a run, it periodically received information from the main DAQ program (see Sec. 4.4.6) such as delays for each channel, which were necessary for a proper alignment of the 192-ns window. Each of the four CPUs ran the FADC-DAQ program separately and wrote the raw data to their local disk. As soon as a run had finished, the data from all four crates was exported and merged with the other data.

4.4.6 Data acquisition system

STACEE used a multi-threaded data acquisition (*DAQ*) program, called `vmedaq`, which ran on a PC unit in the VME crate under *LynxOS*, a POSIX-compliant operating system compatible with Unix/Linux. The CPU used a branch bus to communicate with the rest of the VME crate as well as an adjacent CAMAC crate. The DAQ program had a multitude of functions, the most important of which were: user control of the data acquisition, arming and disarming the system for accepting the different types of triggers, calculating and setting of the channel delays, reading out all relevant electronics modules, inhibiting (or *veto*ing) the system during module readout, and writing of the raw data to disk. Run parameters such as desired trigger type, L1 and L2 multiplicity conditions, run length, discriminator thresholds, ID code of the observed source *etc.* were entered through a text-based interface running on a terminal in the control room, where most of the readout electronics were located. The system allowed the buffering of up to 2000 events during high-rate bursts. During data acquisition, it periodically sent relevant status information to a monitor program running on a separate Linux PC, which displayed important quantities pertaining to the ongoing run such as L1 and L2 rates, PMT anode currents, average discriminator rates *etc.*. in quasi real time. Over the local network, `vmedaq` also periodically communicated with the DAQ programs running on the four FADC crates, providing the latter with important information such as start- and end-of-run signals as well as the current delays for all channels (see Sec. 4.4.5). All five DAQ programs, *i.e.* `vmedaq` and the four FADC DAQs, wrote their data to a local disk first. When a run or run-pair had finished, the VME and FADC-based data were imported via a script run on a separate Linux PC, where they were subsequently merged by examining the time stamps of individual events in all five files. Since the heliostat and weather data were not available until the next day (see Sec. 4.3), these were only merged with the VME and FADC data later.

Apart from the data from event triggers, `vmedaq` also generated so-called *environment triggers* once every 30 seconds, used to read out the PMT anode currents from the Scan-ADCs as well as all scalers. These data were also written to disk, along with a time stamp.

The remote PMT high voltage control system was completely separated from the DAQ, as already mentioned in Sec. 4.4.2. However, `vmedaq` had the ability to shut down the high voltage if it detected PMT currents in excess of some user-tunable threshold, which was set to 100 μA for data taken in 2003 and 2004.

4.5 Electronics calibration and monitoring

The STACEE electronics boasted a multitude of tunable parameters upon which the proper functioning of the detector depended quite sensitively. Furthermore, any meaningful reconstruction and analysis of the recorded data relies on the fact that the quantities in the data stream were calibrated, or at least can be calibrated offline by using the information at hand.

While the description of the calibration process of each and every element of the STACEE electronics is well beyond the scope of this thesis, several quantities are important enough for the result of any data analysis that they deserve to be examined here in some more detail. Since the laser calibration system plays a crucial role in determining important parameters such as the PMT gains and the timing resolution, the first part of this section is devoted to a brief description of this equipment. Then we describe the methods used to measure and calibrate those quantities that ultimately determine the energy response of the detector, namely PMT gains and discriminator thresholds.

4.5.1 Laser calibration system

While some information on the status of the various electronic systems could be obtained simply by examining "ordinary" Cherenkov data, measuring quantities such as PMT gains, timing resolution and trims required special runs taken under certain controlled and well-defined conditions. In order to measure and monitor quantities such as the PMT gains and the timing resolution, STACEE used a laser calibration system, the details of which are described in Hanna & Mukherjee (2002). In the framework of this work only the most important aspects shall be pointed out. A block diagram of the laser calibration system can be seen in Fig. 4.27. It was based on a 100 μJ pulsed nitrogen laser (*Laser Science VSL-337ND-S*), a stilbene dye module (*Laser Science DLMS-210*),

a series of adjustable neutral density filters and a network of optical fibres. The laser was controlled by TTL signals from the DAQ. It emitted ~ 4 ns wide pulses of 337 nm UV-light to the dye module, which converted it into blue light of wavelength 420 nm. The pulse frequency was variable, but set to 10 Hz during normal usage.

Downstream of the dye module, a beam splitter split the light pulses into one part travelling towards a photodiode, which was used as a trigger by discriminating its output voltage and feeding it into the trigger circuit, and another part travelling towards a series of neutral density filters. The neutral density filters contained both discrete and continuous filter wheels, whose attenuation changed logarithmically; two discrete filter wheels could only be adjusted manually, whereas one continuous filter wheel could be used either in static mode or automatically turned by a stepper motor using TTL pulses from the DAQ. Together, the three layers of filters had an attenuation range of over three orders of magnitude. After having passed through the filters, the pulses were fed into a series of optical fibres leading up to the camera boxes on the 220 and 260 ft. levels. They ended in a wavelength shifting plastic located at the top of the cameras, reradiating the light at 500 nm. All of the PMTs in the five cameras were illuminated simultaneously with the fast laser pulse, simulating a Cherenkov wavefront to reasonable accuracy. Another fibre ran to the inside of a DTIRC lens cover, and could be used as a controlled source of light pulses for measuring the PMT gains.

Furthermore, some of the light that had passed through the filter wheels was branched off and directed at a number of photodiodes. The latter were used to monitor the laser intensity, which fluctuated up to 10% between shots. The photodiode charges were measured through charge-integrating ADCs (analogue-to-digital converters). The ratio of the measured pulse charges on any two photodiodes showed an RMS jitter of less than 1%, indicating that the photodiodes themselves were very stable. The relative timing of pulses between different photodiodes had an RMS of 0.3 ns.

The laser system could also be used without the photodiode trigger. In this case, the DAQ enabled the normal trigger system (MADDOG), but instead of applying the TOF delays for Cherenkov observations, it set the relative delays to the differences in light propagation time between the laser flashers and individual PMTs, which were in the order of a few nanoseconds. Taking laser data with the MADDOG trigger was advantageous in several situations, one application being the measurement of the total propagation time of the Cherenkov signal through the electronics (*i.e.* the trims).

The measurement of relative and absolute gains was undoubtedly the most important application of the laser calibration system. Details of both methods will be described in the following section.

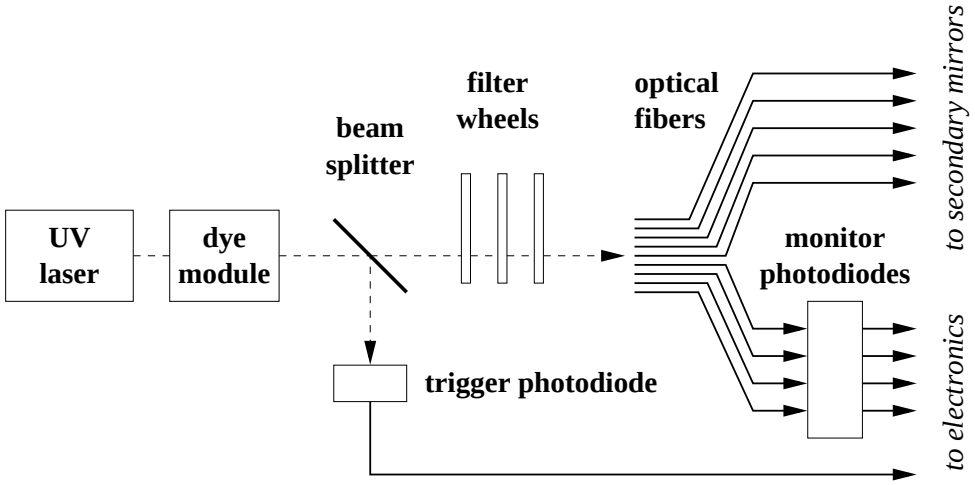

Figure 4.27: Block diagram of the laser calibration system, adopted from Hanna & Mukherjee (2002).

4.5.2 Gain measurement

The gains of the individual photomultiplier tubes, defined as the average number of electrons in the PMT anode current for each photoelectron released from the photocathode, were the quantities upon which the response of the complete detector depended most sensitively. STACEE's trigger rate, and therefore sensitivity to Cherenkov light, increased with the gains in a roughly exponential fashion. Therefore, small changes in the gains resulted in large changes of the detector sensitivity. Moreover, as will be described later in Chapter 5, relatively small errors in our estimate of the PMT gains can result in a substantial under- or overestimation of the detector sensitivity and energy threshold. Therefore, measuring and monitoring the gains was of critical importance. Usually, due to the more complicated procedures required, *absolute* gains were measured only periodically, while *relative* gains could be extracted from calibration data taken every observing night.

Relative gain measurements using photodiode charge

By means of the monitor photodiodes it was possible to obtain a measurement of the relative PMT gains, which manifested themselves in the ratios of the PMT pulse charges to the photodiode pulse charges. In order to remove any charge-related biases and to test the linearity of the method, a range of pulse heights was sampled by using the turning filter wheel driven by the stepper motor. This procedure was used for monitoring the

PMT gains on a nightly basis. Changes in the charge ratio of a channel indicated a change in gain, indicating a possible problem with the PMT or the electronics.

Absolute gains via photoelectron statistics

Several different ways of measuring the absolute PMT gains were implemented at STACEE. The most robust (although somewhat time consuming) method made use of photoelectron statistics and integrated charge as measured with the FADCs to deduce the charge of a pulse resulting from a single photoelectron. This procedure, also called *FADC Gain Inference Technique* (FGIT), is described in detail in Fortin (2002), and shall be summarised in the following.

A special PMT lens cap, which had a small hole in the centre through which one of the fibres from the laser was led so that it illuminated the inside of the cap, was placed on the PMT can whose gain was to be measured. The laser filter wheels were adjusted so that the intensity of the laser light received by the PMT was reduced to an average of between 0.2 and 1.0 photoelectrons per laser shot. Data was then acquired with the laser set to the maximum stable rate (10 Hz) using the photodiode trigger. In the analysis, the FADC waveforms were integrated within a small (20 ns) window around the expected time of the laser pulse, determined by the (known) arrival time of the photodiode trigger. A window of identical width, located in the background region of the FADC trace about 70 ns before the signal window, was also integrated. The charges from both signal and background regions were histogrammed, and the background histogram was fit by a Gaussian distribution. Since there should be no contribution from laser light in this distribution, the obtained Gaussian corresponded to the charge pedestal, or charge distribution for zero photoelectrons. Similarly, the charge histogram for the signal region consisted of a superposition of Gaussian distributions with means corresponding to the charges due to 0, 1, 2, 3 *etc.* photoelectrons as determined by the Poisson distribution, which predicts that the probability of obtaining n photoelectrons is

$$P(n) \;=\; \frac{\mu^n \, e^{-\mu}}{n!} \tag{4.2}$$

where μ is the mean number of photoelectrons produced by a single laser shot. The location and width of the pedestal Gaussian was used to constrain the corresponding zero-photoelectron peak in the charge distribution of the signal region. Furthermore, from 4.2 it follows that the normalisations of the two Gaussians, $N_{0,bg}$ and $N_{0,sig}$, are related via

$$\frac{N_{0,sig}}{N_{0,bg}} \;=\; e^{-\mu} \;=\; P(0) \tag{4.3}$$

Using the location and width of the zero-PE peak as well as the relation between μ and $N_{0,sig}$ as constraints, the charge histogram of the signal region was then fit by four different Gaussians, making further use of the fact that the charge values for the $n = 2$ and $n = 3$ photoelectron peaks had to be integer multiples of the peak corresponding to a single photoelectron ($n = 1$). (Contributions of terms with $n > 3$ were negligible, as adding the corresponding Gaussians did not improve the fits.) The gain of the whole system, consisting of the PMT gain and the gains of the amplifiers as well as the attenuation caused by the long signal cables, was then simply the charge at which the Gaussian for one photoelectron peaked divided by the elementary charge e. Amplifier gain and cable attenuation had been measured independently and were found to be 98.0 and 0.88, respectively, so the "bare" PMT gains could readily be obtained by dividing the total gains by 86. Fig. 4.28 shows a typical example of superimposed signal and background charge histograms.

While the FGIT method provided the best estimate for the PMT gains, it required separate laser runs for each of the 64 channels, and was thus quite time consuming and tedious to perform. In practicality, the gains were measured via FGIT only about once every observing season for all channels. The gains obtained through FGIT were then used to calibrate the relative gain measurements described in Sections 4.5.2 and 4.5.2, which were done on a nightly basis using the standardised 3-minute laser runs or regular Cherenkov data, respectively.

Gain measurements using the FADC trace variance

For completeness' sake, a third method of measuring the PMT gains shall be described here briefly. It is quite intuitive that the gain of a PMT directly affected the variance of the signal due to night-sky background, *i.e.* the FADC trace in the absence of any Cherenkov light: the larger the gain, the larger the amplitude (or RMS) of fluctuations due to NSB photons for a certain constant NSB rate. The anode current, on the other hand, was directly proportional to the NSB rate. If G is the gain, R the rate of NSB photons actually producing a photoelectron, e the electron charge and $I = GeR$ the PMT anode current, a simple derivation yields

$$\frac{\sigma^2}{I} = \frac{GeZ^2}{\Delta t} \tag{4.4}$$

where σ is the variance of the FADC trace, $Z = 50\,\Omega$ the impedance of the signal cable and Δt the FWHM of a typical single-photoelectron pulse in the trace. Eq. 4.4 thus confirms that the ratio of the square of the trace variance to the PMT current is proportional

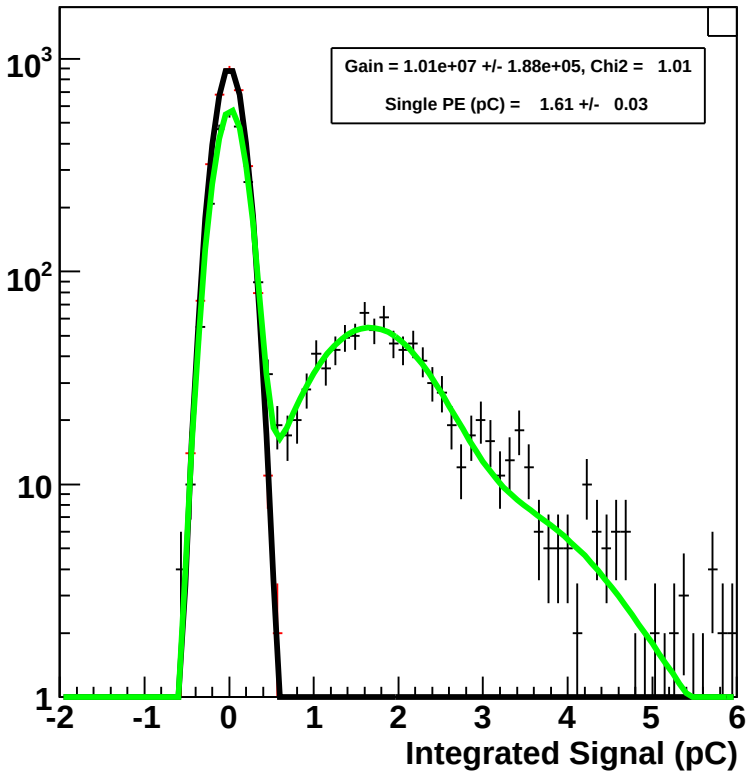

Figure 4.28: Typical laser pulse charge distributions for low light levels, as used when measuring PMT gains (*FGIT* procedure). The charges are obtained by integrating the FADC traces in a window around the expected laser signals (*black points*). The result of the fit is shown in green; the peaks due to zero, one and two photoelectrons can be identified. The single Gaussian fit to the integrated charges of the background (pedestal) window is shown in black.

to the gain. The gain of a channel could then easily be obtained by plotting σ^2 versus I and applying an origin-constrained linear fit to the data. Since both relevant quantities are readily accessible in the data stream for all runs, a gain calculation "on the go" was in principle possible, allowing both measurement and monitoring of the absolute gains without the need for special runs. Unfortunately however, the proportionality constant is not known well enough to calculate reliable absolute gains using this method; the effective widths Δt of the single-photoelectron pulses varied slightly between individual channels and were difficult to measure, and hence not known to sufficient accuracy. Furthermore, the method suffered from systematic effects caused by the high levels of RF noise in some channels, which spoiled the linearity of σ^2 with respect to I by adding a constant offset to the former variable. For these reasons, the trace variance method was

used only to calculate relative gains, and combined with the FGIT technique, provided a very convenient way of monitoring the gains on a daily basis.

Gain adjustments

It was found that the PMT gains decreased over the course of an observing period, most likely due to photocathode wear caused by the repeated long periods of exposure to high background light levels. If the gain of a PMT needed to be adjusted, the applied high voltage was varied in steps of 50 Volts around the previous high voltage value. At each setting, a laser run was taken and relative gains were measured, which were subsequently plotted as a function of high voltage. The empirical relationship between gain G and PMT high voltage V,

$$G \propto V^{\gamma} \tag{4.5}$$

where $\gamma > 1$, allowed an exponential fit to the data. The fitted curve could then be used to find the required high voltage that would yield the desired relative gain.

4.5.3 Timing calibration

As explained earlier (see *e.g.* Sec. 3.1), the Cherenkov light wavefront of an EAS is only a few light-nanoseconds thin. Knowing the exact *timing resolution* of the detector, corresponding to the accuracy to which the arrival times of Cherenkov light pulses could be recorded, is important for a variety of reasons. It needs to be known to at least 10% accuracy in order to determine the likelihood that an event which had passed the hardware trigger is associated with a Cherenkov light wavefront, and it determines the goodness of the fit of the recorded photon arrival times to a gamma-ray shower profile. Physically, the arrival time of a "hit" on a channel is determined by the time at which the discriminator fired, *i.e.* at which the input waveform exceeded the discriminator threshold (modulo discriminator hysteresis effects, which are not important in this context). If one assumes that the delays of all other components of the electronics chain, *i.e.* the trims, were constant over time, the timing resolution of a channel was dominated by the random oscillation, or *jitter*, of the discriminator firing times around a mean value corresponding to the actual arrival time of a photon at the heliostats.

Under these assumptions, there are two factors on which the timing resolution of a given channel depended: pulse height and noise level. For pulses with large amplitudes, which have steeper leading edges, the threshold crossing time was less likely to be affected by random baseline fluctuations. On the other hand, the effect of noise on the trigger times of small pulses with smaller voltage gradients was substantial. For effectively the

same reasons, the jitter increased with increasing noise level in the analog signal, thus the timing resolution of a particular channel degraded with increasing level of night-sky background experienced by the associated PMT. Likewise, RF noise affected the timing resolution adversely.

In order to provide the fitting algorithms in the offline analysis with an appropriate value for the timing resolution, this quantity needed to be measured for a variety of pulse heights and under a range of different background light conditions. Special runs using the laser in photodiode trigger mode, for which the arrival time of the laser pulses is well known, were used to build up histograms of the discriminator times and fit them with a normal distribution. The widths of these Gaussians then corresponded to the timing errors. A broad range of pulse heights was generated by using the stepper motor to rotate the continuous filter wheel. The procedure was repeated for a set of different background light levels, which were achieved by means of a white LED in the special laser lens cap illuminating the DTIRC, or by placing an LED on the optical axis of the secondary mirror.[5]

Tightly correlated with the timing resolution is another relevant characteristic of a pulse, namely the dependence of the discriminator firing time on the pulse height, also known as *slewing*. Using the same laser data as above, the FADC traces were analysed and the time of discriminator firing as a function of pulse amplitude plotted. An empirical fit to these data was made, and the fit parameters are used in the offline analysis to correct all arrival times of triggered events for pulse slewing. Fig. 4.29 shows an example of a slewing curve.

4.5.4 Discriminator thresholds

The ultimate goal of STACEE was to be sensitive to gamma rays with energies of 100 GeV or less. Achieving such low energy thresholds required the PMT (discriminator) thresholds to be set as low as possible, yet still high enough to avoid accidental triggers due to random coincidences from night-sky background noise.[6] For a simple one-level trigger logic requiring n channels out of a group of m channels to produce

[5] In fact, although the *mean* arrival times of the laser pulses were well defined due to the use of the photodiode trigger, the photons within a single laser pulse were not emitted at exactly the same time, but rather within a roughly uniform time window of about 5 ns width. Therefore, this laser pulse width, assumed to be Gaussian as well, needed to be subtracted from the total widths.

[6] In the absence of any kind of background light or electronic noise, the thresholds could theoretically be set arbitrarily low and the detector would still trigger only on Cherenkov light showers, providing for an astronomically low energy threshold only limited by the double-pulse resolution of the discriminators. In reality, of course, such conditions can never be achieved.

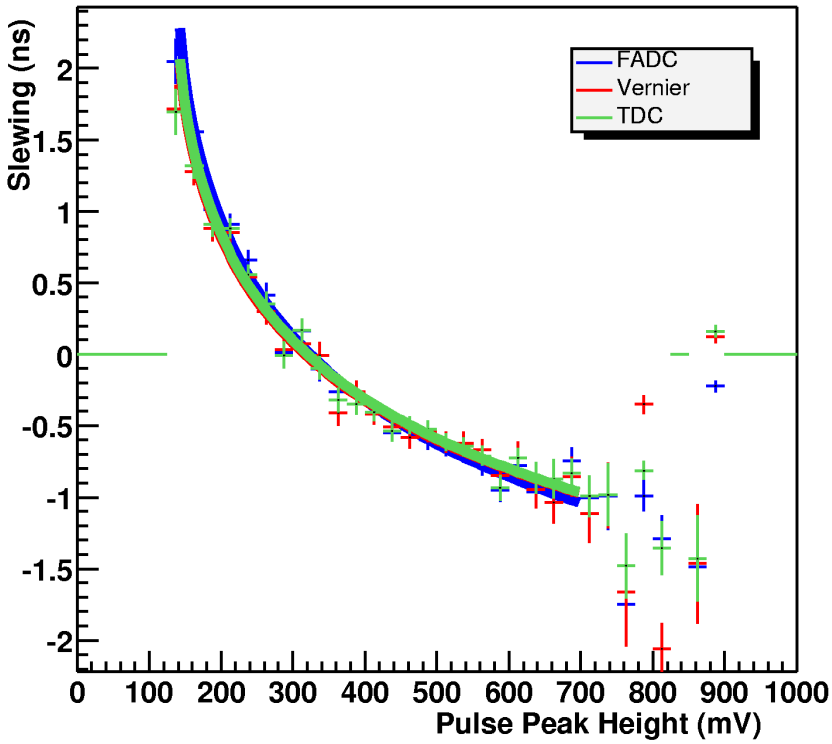

Figure 4.29: Typical slewing curve, *i.e.* the relative discriminator firing time in dependence of the laser pulse height. The location of the zero-point on the time axis is somewhat arbitrary, as only relative times are of interest. Shown are slewing curves obtained by using vernier (MADDOG), FADC and TDC times.

a discriminator output, the expected rate of accidental triggers ρ as a function of the average discriminator rates R and the coincidence width τ can be written as

$$\rho = n \binom{m}{n} R\,(R\tau)^{n-1} \tag{4.6}$$

For any given channel, R increased exponentially with decreasing discriminator threshold. Note that Eq. 4.6 can readily be generalised to the case of a two-level trigger by recursion, *i.e.* by substituting ρ for R. The set discriminator thresholds where the detector was starting to trigger on night sky background noise, usually defined as the common threshold at which the accidental coincidence rate is reduced to 10^{-2} Hz or less, is called the *breakpoint*. The location of this point depended naturally on the amount of night-sky background light received by the PMTs, but as indicated by Eq. 4.6 also on the chosen trigger multiplicity conditions (LTC and GTC) and the width of the trigger coincidence gate. In theory, setting the discriminator thresholds to the value of the breakpoint would add a negligibly small contribution of accidental triggers to the real triggers from

Cherenkov showers. During normal operation of the detector, however, factors such as changes in the regions of the sky viewed by the heliostats, changing overall sky conditions or slight gain drifts due to things such as temperature changes in the electronics all affected the location of the breakpoint. Therefore, the discriminator thresholds were usually set to values about 10 mV *higher* than the breakpoint. The location of the breakpoint was determined before observations of any given source by measuring the trigger rate as a function of discriminator threshold while tracking a point in the same sky region as the potential gamma-ray source. The rate at each point was measured twice: once with properly timed in delays and once with random ("*scrambled*") delays, yielding curves such as the ones shown in Fig. 4.30. The rates with random delays were due to accidental coincidences alone, as they contained practically no contribution from actual Cherenkov showers, and depended only on the background currents, LTC and GTC multiplicities, and trigger coincidence width. For the data used in this work, the discriminator thresholds had been set to values between 115 and 140 mV.

Threshold measurement

Finally, the stability of the discriminator thresholds needed to be monitored. Given that the latter were responsible for the energy threshold of the detector, measuring the actual, or *effective* thresholds was an essential step on the way to interpreting the data. To that aim, the following procedure was implemented: for each ordinary 28-minute run, the FADC waveforms were discriminated in software at various thresholds V_i, covering the physically sensible voltage range in sufficiently fine steps. Since every pulse that contributed to the MADDOG trigger had a vernier hit associated with it (see Sec. 4.4.4), and the same pulses had been digitised and written to disk, all vernier hits must have a counterpart in the FADC data. A one-to-one correspondence between hits was asserted by requiring a coincident vernier hit within a few ns of the FADC hit time, defined as the time at which the FADC waveform crossed the given discriminator threshold V_i after subtraction of the trims and all FADC offsets. If there was no coincident FADC hit found at threshold V_i for a given vernier hit, the actual hardware threshold must have been *lower* than V_i. Conversely, FADC hits with no coincident vernier hits indicated that the hardware threshold had been *higher* than V_i.

Fig. 4.31 shows the fraction of triggered events which had coincident FADC and vernier hits, as a function of software threshold imposed on the FADC waveform, for a representative run. A probability function of the form

$$P(V) = a + b \ \mathrm{erf}\left(\frac{V - V_0}{\sigma_v}\right) \tag{4.7}$$

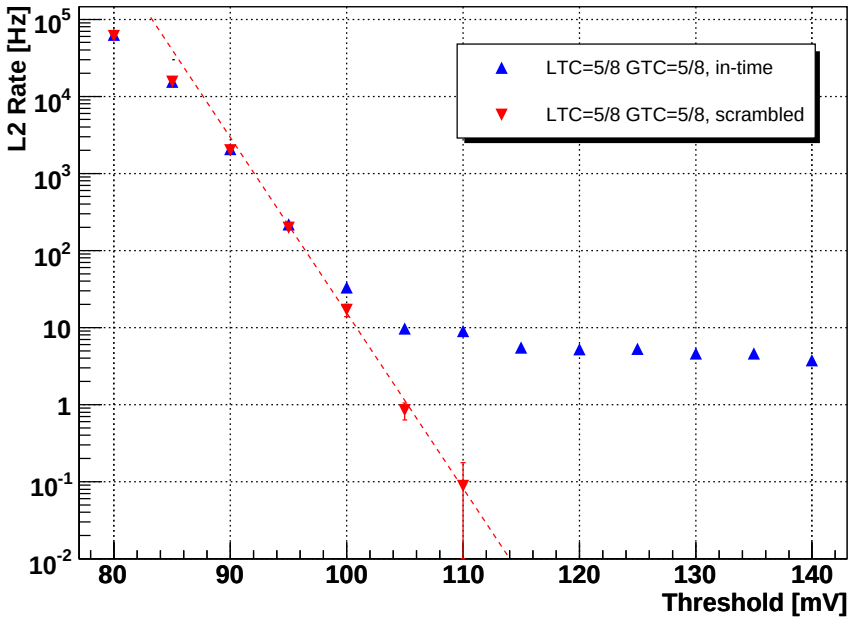

Figure 4.30: L2 trigger rate for hadronic air showers as a function of discriminator threshold, measured in April 2004. The trigger conditions were GTC=5/8 and LTC=5/8 with wide trigger mode. Blue triangles represent the rate with proper delays for gamma-ray observations. The contribution from accidental coincidences, measured by randomising the applied delays, is plotted as red inverted triangles. The dotted line is an exponential fit to the accidentals rate, corrected for the saturation due to DAQ deadtime visible at low thresholds. The intersection of the exponential with the 10^{-2} Hz axis defines the so-called *breakpoint* (see text), which for this plot is at 114 mV.

was fitted to the distribution. The turning point of the Error function (*erf*), which is where the hardware trigger was 50% efficient, corresponds to the average (or effective) threshold of that discriminator channel. The gradual turn-on, reflected in a rather large σ_v parameter, can be attributed to several factors: a systematic error arising from the fact that the FADC values were quantised due to the digitiser's rather coarse voltage resolution of 3.9 mV, intrinsic effects of the discriminators such as hysteresis and the requirement of a finite time-over-threshold, and lastly the finite double-pulse resolution of the vernier codes. It is likely that the latter is also responsible for the fact that the efficiency curve plateaus out below unity.

For reasons which were never fully understood, the effective thresholds showed relatively large deviations from the set thresholds. As an example, Fig. 4.32 shows the measured thresholds for all of the data analysed in this work that were recorded with a nominal threshold of 130 mV. The most striking feature of the plot is that the 48 chan-

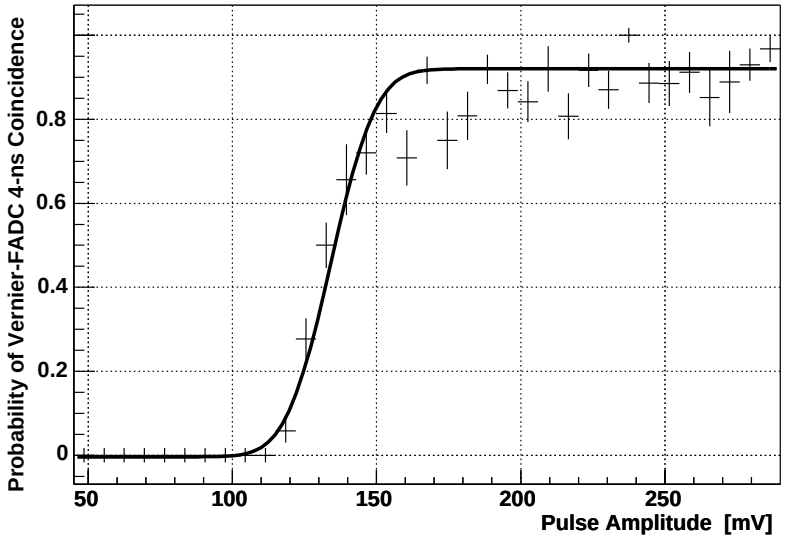

Figure 4.31: Measurement of a channel's effective discriminator threshold: the trigger probability, defined as the fraction of triggered events having coincident hits in the FADCs and the verniers, is plotted against software threshold and fitted with an *erf* function, whose turning point is taken to be the effective threshold (here ≈ 135 mV).

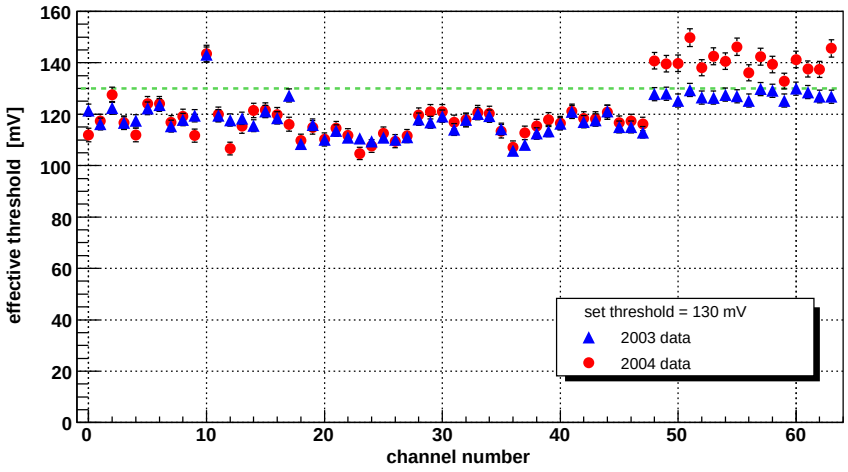

Figure 4.32: Average measured effective discriminator thresholds for the data on H 1426+428 taken with a set (nominal) threshold of 130 mV, indicated by the dashed line. The errors shown are approximate. Blue triangles represent data taken in 2003, red circles data from 2004. In both data sets the discrepancy between the South camera channels (channels 48 - 63) and the rest is clearly visible, although it is significantly less pronounced in the earlier data. Note that except for a few exceptions, the thresholds of the 260-level channels did not change much in between the two observing periods.

nels associated with the 260-level cameras seemed to trigger at thresholds 5 to 25 mV *lower* than the set threshold, whereas the South camera channels had effective thresholds of just below nominal and in average 10 mV *higher* than nominal for the 2003 and 2004 data, respectively. There is some consensus that part of this different behaviour could be attributed to the different kinds of discriminator modules employed within the two groups of channels (LeCroy 4413 and Phillips 7106, respectively). Instead of the nominal (set) thresholds, the effective thresholds are used in both data analysis and simulations.

CHAPTER 5

SIMULATIONS

When speaking of simulations in the context of gamma-ray astrophysics – just as in nuclear and particle physics – we commonly mean *Monte-Carlo* simulations. As suggested by the name, Monte-Carlo techniques are based on random number generation. Given that physical quantities such as reaction cross-sections or the exact locations and momentum distributions of particles within a medium are based on probabilities, so must any modelling of such quantities. In principle, the Monte-Carlo technique thus consists of choosing initial conditions – either well defined or at random, folded with whatever applicable probability distribution – and letting them evolve through a given model, which itself usually consists of a long sequence of random processes. It is already clear from this concept that in order to obtain a physically meaningful answer, *e.g.* one which can be compared with observable quantities, this process needs to be repeated a large number of times. The more possible initial and final states there are, the more trials are required. The statistical errors incurred decrease with the number of trials, whereas systematic errors, which result from the unavoidable imperfection of a given model, remain constant.

The proper understanding of any detector that is sensitive to highly energetic particles requires knowledge about the large variety of possible physical processes inside the detector volume, and how they map a given set of initial conditions to a set of observable quantities. In the real detector, these initial conditions are not always known, nor is there a way for most of the intermediate processes to be measured directly. However, provided the exact composition of the detector is known in sufficient detail and also the reaction cross-sections for all relevant types of particles in the applicable kinematic range, Monte-Carlo simulations can provide the missing link; for in "simulation land", the initial conditions are well defined, and all intermediate processes are known. In nuclear and particle physics, therefore, Monte-Carlo type simulations are absolutely indispensable for understanding the detector and proper interpretation of the measured quantities.

Other cases in which simulations are necessary include situations in which the intermediate processes are in principle measurable by examining one element of the detector

at a time, but contain elements that are random in nature, and/or contain an exceedingly large number of "nodes" and thus possible outcomes. Such situations prohibit the use of look-up tables or similar straight-forward parameterizations, and require the detector to be modelled.

In the case of ground-based gamma-ray (and also cosmic-ray) detectors, which cannot be calibrated by means of a well defined test beam of gamma rays (or cosmic rays), the importance of proper detector simulations is particularly evident.[1] For converting an observed signal, or lack thereof, to an actual flux or upper limit, the detector response needs to be know as a function of particle energy and angle of incidence. Often, sources of background and how they affect the detector need to be known as well. In the case of ground based gamma-ray telescopes, any observational results have thus little scientific value without a proper simulation of the detector.

It is immediately clear that the task at hand is not an easy one. For atmospheric Cherenkov telescopes, the physical detector volume includes the Earth's atmosphere – a medium whose physical composition and properties change with altitude and also with time. Models of this category of detectors are therefore inherently complex and also prone to systematic errors. Fortunately, there still exist ways to "calibrate" the simulations, by means of which systematic errors can be minimised. In the following we will describe the set of simulation programs used to model the STACEE gamma-ray telescope.

5.1 The STACEE simulation chain

For STACEE, and in fact for any ACT, the "detector" to be modelled consists of three separate and well defined components:

1. Atmosphere
2. Optics
3. Electronics

Owing to this physical division, STACEE's complete detector simulation comprises three separate programs, together also referred to as the *simulation chain*, each responsible for modelling one of the above "layers".

[1] The object coming closest to a "test beam" for gamma-ray astronomy is the Crab Nebula, but for calibration purposes even the flux from this object is extremely low, and its emission spectrum and position in the sky cannot be varied as needed. Last but not least, the gamma-ray flux from the Crab is not yet known well enough over all energy ranges.

The task of simulating an extended air shower is rather complex, in particular when hadronic processes play an important role. Like most other ground based gamma-ray and cosmic-ray experiments, STACEE therefore uses one of the publicly available air shower simulation codes to simulate air showers and the production of atmospheric Cherenkov light. Initially this was the `MOCCA` code (Hillas, 1995), but later that software was replaced by the more detailed `CORSIKA` (Heck *et al.*, 1998; www-CORSIKA, 2008). The output routines of both air shower simulations were written by the STACEE collaboration and are linked with the program during compilation. They ignore all charged particles that are traced to ground level, and only write out the information pertaining to the produced Cherenkov photons, traced down to the observation level, in a compressed format. These files are then read in by the next stage of the simulation.

The Monte-Carlo simulation program that models the STACEE optics is called `Sandfield` and was written by the collaboration (see *e.g.* Oser, 2000). This program is essentially a wrapper for a ray-trace, reading in lists of Cherenkov photons from the output of CORSIKA and tracing each of them through the primary and secondary optics onto the front surfaces of the DTIRCs. Whenever a photon hits a DTIRC, the program consults a look-up table in order to determine the probability of that photon releasing a photo-electron from the photo-cathode. The output of this program is thus a list of photoelectrons for each PMT, along with associated time stamps.

The final stage of the STACEE simulation chain is the electronics simulation. In the early days of STACEE, no detailed modelling of the experiment electronics was done, and simulations were concluded with routines that performed a crude counting of photoelectrons as output by the optical simulation, and employed simplistic trigger models. This usually resulted in a grave underestimation of the detector's energy threshold. The first "true" electronics simulation, `HERMES` (Boone, 2000), was written for the STACEE-48 prototype detector and contained a rather detailed model of the discriminators and trigger. When STACEE was upgraded for the use of 64 heliostats, `HERMES` was replaced by a simulation program with enhanced performance called `Elec`, which subsequently became the standard electronics simulation for STACEE.

Apart from the three programs, various driver scripts were written by the collaboration; these allow a user to run two or even all three stages of the simulation together in one pass. Most often, however, the electronics simulation is run separately, for reasons that will be explained later. For interpreting simulation results and assessing sources of systematic error it is important to understand the basics of how the simulation algorithms work, and what assumptions go into the detector Monte-Carlo. In the following sections, the three simulation programs will therefore be examined in greater detail.

5.2 Air shower simulation: *CORSIKA*

Over the last decade, the CORSIKA (*COsmic Ray SImulations for KAscade*) air-shower simulation program has become the *de facto* standard in cosmic-ray and VHE gamma-ray astrophysics. It contains a more detailed treatment of many processes and a more modular approach in comparison to some earlier simulations, such as MOCCA and *AIRES*. The modularity allows for the usage of detailed, pre-existing codes widely used in particle physics, such as *EGS 4* for electromagnetic interactions, three different simulation codes for low-energy ($\lesssim$ 100 GeV) and six for high-energy ($\gtrsim$ 100 GeV) hadronic interactions. A very large number of primary particle types are supported, including many mesons and baryons as well as nuclei up to $A = 59$. Due to its widespread use in the high-energy astrophysics community, CORSIKA has been adapted to suit the needs of a variety of different experiments that – directly or indirectly – are sensitive to air shower particles or their by-products[2], such as AMANDA, AUGER, HEGRA, MAGIC and, of course, STACEE. As such, it supports many different detector configurations and particle output options.

CORSIKA consists of four basic modules. The first is the actual program frame which calls the various specialised subroutines, controls in- and output, handles the decay of unstable particles, and the tracking of particles in the geomagnetic field while taking into account ionisation energy loss and deflection by multiple scattering. The second and third parts are essentially stand-alone subprograms that calculate the details of the hadronic interactions at lower and higher energies, respectively. The particle energy at which the program switches between the two packages is user-tunable, and generally in the range of 80 - 100 GeV. The fourth part describes the transport and interactions of electrons, positrons and photons, *i.e.* the electromagnetic shower component. In Sections 5.2.1 and 5.2.2 we shall give an overview of the various particle interaction codes.

The simulation concept utilised by CORSIKA is straightforward. After determining the atmospheric depth and atmospheric target particle of the first interaction, cross-sections for possible reaction channels are evaluated, a suitable set of secondary particles and their momenta are chosen, and the developing air shower is traced through the atmosphere by iteratively treating all secondary particles in the same manner, until the particles' energies fall below a (particle type specific and user adjustable) energy threshold. Optionally, to avoid excessively long processing times for the large showers of very energetic primaries, a so-called *thinning* algorithm is invoked when the energy

[2]Mainly Cherenkov and air-fluorescence radiation.

of a shower particle falls below a given fraction of the primary energy.[3] If a corresponding option is selected, CORSIKA also traces all Cherenkov photons generated by the shower particles in the atmosphere, taking into account atmospheric attenuation and refraction. The STACEE collaboration has developed its own Cherenkov photon output format, implemented in a separate routine written in `C` and linked with the main program (see Sec. 5.2.4). The external interaction routines are universal particle-physics codes, developed for the high-energy physics community, and are written in a variety of programming languages (mostly `C`), while CORSIKA itself is coded in `Fortran`.

5.2.1 Hadronic shower component

While the electromagnetic and weak interactions can be calculated with reasonable accuracy up to very high energies, cross-sections of the strong processes (and thus of hadronic interactions) are still affected by large uncertainties. Currently results from $p\bar{p}$ colliders and heavy-ion accelerators provide for cross-section measurements at centre-of-mass energies up to the TeV range, corresponding to proton energies of $E \gtrsim 10^{17}$ eV in the lab frame. Apart from large errors due to the limited statistics, the problem with current data is that most of it is applicable to large momentum transfers, *i.e.* transverse scattering in the CM-system. This is a consequence of the fact that particle detectors of the highest-energy colliders (such as the *Tevatron* at Fermilab) rarely register events in the forward directions, since there is no detector coverage at solid angles close to the beamline. The kinematic region of small momentum transfers however is the one that is most important for air shower physics, since these types of interactions give rise to the most energetic daughter particles, and hence dominate the strongly beamed part of an EAS. This lack of data is dealt with by extrapolating the available measurements to even higher energies and smaller angles; it is therefore not surprising that different hadronic models may be implementing different interpretations of the sparse data.

As mentioned earlier, its modular structure allows CORSIKA to be linked with a variety of different hadronic interaction models. At the time when the last STACEE version of CORSIKA was compiled (Sec. 5.2.4), there were three packages available for covering the energy range up to $\sim$100 GeV, and the user had the choice between six different codes for the energy range above that. This section is meant to give a brief overview of the models, particularly in the context of STACEE's simulation needs.

[3] This is of some importance for experiments simulating UHE cosmic rays such as AUGER; for STACEE, where the highest relevant particle energies are in the range of a few TeV and the total number of secondary particles produced is much smaller, activating thinning could introduce unwanted errors.

Low-energy hadronic models

The three different low-energy hadronic interaction codes available for linking with CORSIKA were *GHEISHA 2002, FLUKA (2003)* and *UrQMD*:

- GHEISHA (Fesefeldt *et al.*, 1985), which was the original hadronic interaction model used by *CERN*'s `GEANT` detector Monte-Carlo package, is the fastest running low-energy code. Unfortunately, the original version (*GHEISHA 600*) was found to suffer from deficiencies in the reaction kinematics (see *e.g.* Heck, 2004), and has therefore difficulties in reproducing some of the measured cross-sections and their energy dependence. Due to the known problems of GHEISHA 600, which STACEE has been using by default until 2004, a patch had been developed and applied by a team of programmers from *SLAC*. The fixed version was made public as *"GHEISHA 2002"*, although it turned out that the patch resolved only part of the problems. In spite of this, GHEISHA 2002 was made available for STACEE users, since its execution speed is faster than that of the competing codes, and the known deficiencies are overshadowed by the (probably much larger) overall systematic uncertainties affecting simulations of the STACEE detector.

- *FLUKA* (Fasso *et al.*, 2001; www-FLUKA, 2008) is a particle physics simulation package that relies heavily on look-up tables, making the code, which was only made public in the form of object (*i.e.* pre-compiled) files, quite large in terms of file size. As far as the reproduction of collider data is concerned, and in contrast to GHEISHA, it manages to achieve a satisfactory agreement in all quantities important for air shower physics.

- *UrQMD* is the name of a model that is based on a microscopic treatment of hadronic air shower particles, and is therefore very detailed. The main consequence is that the necessary processing times are very long (several times those of FLUKA). Along with the fact that agreement with experimental data is good, but not better than for the latter, this renders UrQMD rather impractical for simulating air Cherenkov experiments such as STACEE.

The GHEISHA model's difficulty in predicting the measured pion multiplicities can be seen in the upper two plots of Fig. 5.1. The same plots also suggest that the FLUKA (2003) model describes the data well – which is maybe not surprising, given that it is an interpolation of this same data – and that QGSJET 01 overpredicts the multiplicities in the transition region between low- and high-energy hadronic models. The lower plot, although for a primary energy much higher than those relevant for ACT,

gives an idea of the large differences in Cherenkov light yield for the various different choices of hadronic models (note the logarithmic density scale).

The default low-energy hadronic model used by the STACEE collaboration from November 2004 on was FLUKA (2003) (see www-FLUKA, 2008).

High-energy hadronic models

Comparison of the various high-energy hadronic models for CORSIKA is slightly less straightforward than that of the low-energy models, because there is not as much experimental data available which could be used to constrain the models – at least not at the EeV end of the spectrum. However, since these energies are of no interest to STACEE, we will focus on the performance of the models at the low-energy end (100 GeV $\lesssim E \lesssim$ 10 TeV).

The model favoured by high energy experimental data at the time when STACEE's simulations were initially developed was *QGSJET 01* (Kalmykov *et al.*, 1997), and this model has been the standard for STACEE's CORSIKA simulations. Alternative versions of CORSIKA have been compiled using the *SIBYLL 2.1* (Engel *et al.*, 1999) code. This model was chosen because it predicts a relatively high average elasticity in pion-nucleon collisions, as opposed to QGSJET 01, which yields elasticities that are somewhat lower than those of the other models (see lower panel in Fig. 5.2). As a result, air showers simulated with QGSJET decay slightly faster and do not penetrate quite as deeply as with SIBYLL. Using the latter as an alternative high-energy model thus promised to provide a good handle on the systematic error due to the choice of the hadronic model. Differences in proton trigger rates between the two models were however found to be only significant if the models were combined with the GHEISHA 2002 routines. Table 5.1 gives an overview of the results.

CORSIKA Version	Low-Energy hadr. model	High-Energy hadr. model	STACEE proton trigger rate [Hz]
6.016	GHEISHA 600	QGSJET 01	4.50 ± 0.01
6.200	GHEISHA 2002	QGSJET 01	5.10 ± 0.01
6.200	FLUKA (2003)	QGSJET 01	5.25 ± 0.02
6.200	GHEISHA 2002	SIBYLL 2.1	4.61 ± 0.01
6.200	FLUKA (2003)	SIBYLL 2.1	5.34 ± 0.02

Table 5.1: Predicted STACEE trigger rates for proton showers from zenith, using different versions of the CORSIKA program. For CORSIKA v6.200, different choices of low-energy and high-energy hadronic models have been selected.

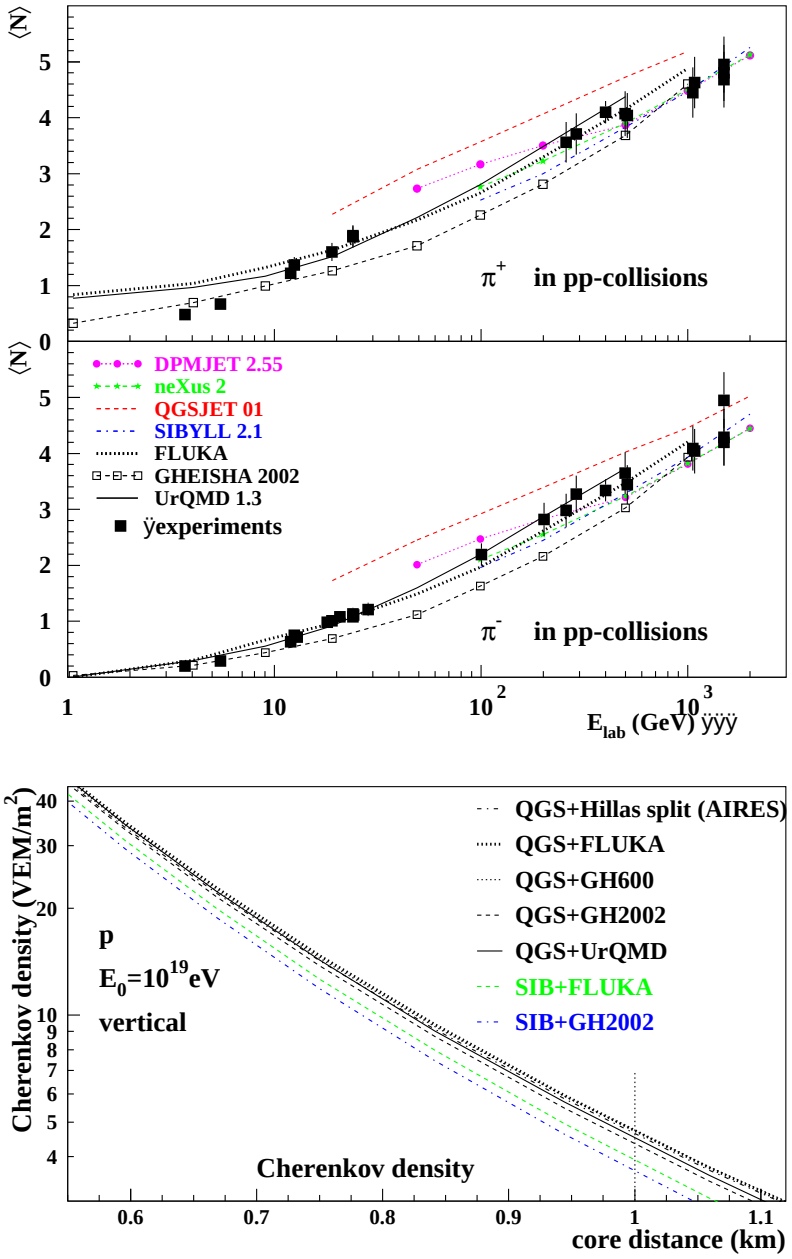

Figure 5.1: Comparison of selected properties of low-energy hadronic interaction models. *Top*: Multiplicities of charged pions from *pp*-collisions as a function of projectile energy, for π^+ (*upper panel*) and π^- (*lower panel*). Shown are the predictions of three low-energy hadronic models (*black lines*) as well as those of selected codes for high-energy interactions (*coloured lines*), and data from collider measurements (*black squares*). *Bottom*: Lateral distributions of Cherenkov photons at large core distances due to a vertical UHE proton shower (energy 10^{19} eV), according to various different pairings of low-energy and high-energy hadronic models. Figures taken from Heck (2004).

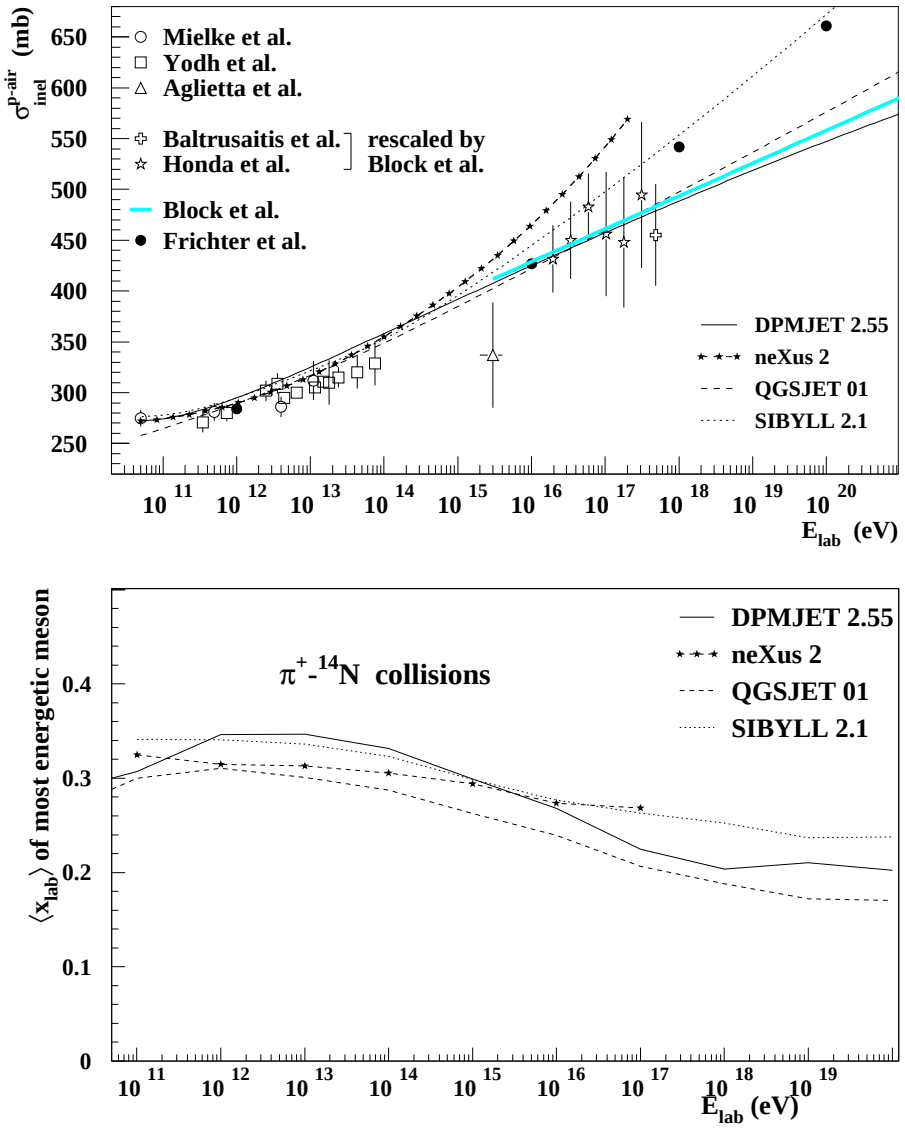

Figure 5.2: Comparison of kinematic properties of high-energy hadronic interaction models. *Top*: Inelastic proton–air cross-sections of selected high-energy hadronic models as a function of proton energy in the lab frame, along with experimental data from air shower measurements. Note the difference between QGSJET and SIBYLL at the very high and the low end of the energy range. *Bottom*: Average of longitudinal momentum fraction carried away by the most energetic meson emerging from π-^{14}N collisions (*elasticity*) as a function of energy, for various high-energy hadronic interaction models. Figures taken from Heck, Risse & Knapp (2003).

Other high-energy hadronic codes available for being linked with CORSIKA are *DPMJET, VENUS, neXus* and *HDPM*. Of these only VENUS appeared to be a promising candidate for usage with STACEE simulations, as the other routines had significant disadvantages either in quality or in execution speed. Recently, a new version of the QGSJET code (QGSJET-II) has become available to be linked with CORSIKA. This model includes non-linear effects in nuclear interactions, and is regarded as an improvement of the original QGSJET. Comparisons of QGSJET-II to other models (Ostapchenko, 2004) however reveal significant discrepancies only at energies well above the STACEE range, and in regard to muon numbers (which STACEE is not affected by).

5.2.2 Electromagnetic shower component

The electromagnetic part of CORSIKA is simulated by the *EGS4* package (Nelson *et al.*, 1985; www-NRC, 2006), which is integrated into the program by default. It handles pair-production and bremsstrahlung, including rare processes such as μ^+/μ^- pair formation, by means of large look-up tables. Since the electromagnetic processes are well understood, so are the physics of electromagnetic shower development, and therefore also Cherenkov light production in gamma-ray initiated air showers (recall Sec. 3.1.1). It goes without saying that this is rather fortunate for ground-based gamma-ray detectors, as large uncertainties in the Monte-Carlo combined with their low signal-to-background ratios would render the ACT technique rather imprecise.

CORSIKA allows to turn on simplifications within EGS4, such as an abbreviated treatment of multiple scattering. Like shower thinning, activating this option is only useful for primaries of extremely high energies, as for smaller shower sizes the quality of the output is seriously degraded. It should be noted that EGS4 runs much faster than any of the hadronic codes, which facilitates the generation of output that requires only gamma-ray simulations, such as gamma-ray effective areas or core templates.

5.2.3 Atmospheric profiles

In CORSIKA, the Earth's atmosphere is modelled as a homogeneous gas with the volume fractions 78.1% N_2, 21.0% O_2 and 0.9% Ar, independent of height (Heck *et al.*, 1998). It is divided into five vertical layers which connect at user-specified boundary heights h_i, typical values of which are $h_{i=1,\ldots,4} = 4, 10, 40$ and 100 km ASL. Within each layer i,

the density, or equivalently the *mass overburden* $T(h)$, is parameterized according to

$$T_i(h) = a_i + b_i \cdot e^{-h/c_i} \qquad i = 1, \ldots, 4 \tag{5.1}$$

$$T_5(h) = a_5 - b_5 \cdot h/c_5 \tag{5.2}$$

The typical exponential decline of the density thus turns into a linear dependence in the outermost layer ($i = 5$), in accordance with measurements of the thermospheric density distribution.

As of CORSIKA version 6.200, a number of internally defined atmospheric parameterizations are supported, as well as the option to read in a user-defined density profile from an external file. The predefined parameterizations are constructed from fits to actual measurements; typically, profiles obtained during the winter months exhibit steeper density gradients, and hence result in the shower maxima being located at slightly lower altitudes than for those representing summer conditions. In general, the Cherenkov light density at ground level is therefore higher for the "winter" atmospheres. Note that atmospheric moisture and aerosol concentrations are *not* taken into account in these standard atmospheric parameterizations. Absolute humidity is lower in colder temperatures, hence the atmosphere is generally more transparent to Cherenkov light in winter than in summer. For many geographic locations, aerosol contents are higher in summer as well (*e.g.* due to forest fires), which also adversely affects the transparency to Cherenkov light.

Differences in Cherenkov yield of typically 15% between winter and summer atmospheric profiles were found by Bernlöhr (2000). The effect of using different atmospheric profiles within simulations of STACEE was investigated by using CORSIKA to generate a large sample of proton-initiated showers from zenith, and process them with the optical and electronics simulation using identical parameters, reflecting typical running conditions. The resulting (L2) trigger rates along with statistical errors are listed in Table 5.2. It is immediately clear that there is a large spread in trigger rates, resulting only from the particular choice of atmospheric parameterization. The standard deviation of this (statistically very limited) sample is just below 10%. The atmospheric profile used for all simulations of STACEE, including those used in this analysis, is the "US standard atmosphere" as parameterized by Linsley (`ATMOD` = 1). We therefore assume a systematic error of 10% on the trigger rate due to the uncertainty in the exact atmospheric parameterization.

ATMOD	Description	STACEE proton rate [Hz]
1	STANDARD (Linsley)	5.254 ± 0.015
17	WINTER I	5.440 ± 0.017
18	WINTER II	4.355 ± 0.015
20	SUMMER	4.770 ± 0.017
22	STANDARD (Keilhauer)	4.935 ± 0.016
EXT. 3	MID-LATITUDE WINTER	4.883 ± 0.016
EXT. 2	MID-LATITUDE SUMMER	4.343 ± 0.027

Table 5.2: Representative simulated STACEE trigger rates for EAS initiated by protons from zenith, with statistical errors, using a variety of different atmospheric parameterizations as available in CORSIKA. *"ATMOD"* is the CORSIKA - ID of the individual parameterizations. Atmospheric profiles read in from external files are labelled *"EXT"*.

5.2.4 The STACEE version of CORSIKA

The CORSIKA source code must be customised during extraction. There are a variety of different options to choose from, pertaining to such things as the type of detector (Cherenkov telescope, particle detector, neutrino detector *etc.*), the detector geometry (single telescope or array, acceptance angle *etc.*) and the energy range to which the detector is sensitive. The hadronic models are also selected at this stage. For STACEE, naturally, the chosen configuration is that of a horizontal, flat Cherenkov detector. Details about the version of CORSIKA customised for STACEE can be found in Mueller (2005); in the following we will point out the most relevant properties.

Since STACEE detects Cherenkov radiation and not charged particles, it is essential that Rayleigh and Mie scattering, as well as absorption, in the atmosphere are taken into account. A customisable data file (`atmabs.dat`) contains tables of atmospheric extinction coefficients as a function of altitude and wavelength, which CORSIKA applies to the Cherenkov photons in the step where they are emitted. Just like the atmospheric model, the values within this file were kept constant for all simulations.[4]

Given the large number of Cherenkov photons produced by energetic air showers, the STACEE version of CORSIKA also supports the application of Cherenkov photon efficiency cuts before the photons are written to disk. This is done in order to save disk space and write access time. The (wavelength dependent) efficiencies are contained in another external text input file (`quanteff.dat`). As hinted at by its name, this file was

[4]Ideally the atmospheric extinction coefficients as well as the atmospheric model should be adapted to the given local observing conditions at the time the data were acquired. Since STACEE lacked a proper tool (such as a LIDAR system) to measure the extinction, averaged values were used throughout the observing season.

originally intended to describe nothing more than the quantum efficiency of the PMTs. For ease of maintenance, all the other *global* wavelength dependent efficiencies of the detector, *i.e.* the heliostat and secondary mirror reflectivities as well as the transmissivity of the DTIRCs, were also accounted for in this file.[5]

Input parameters

In order to generate an air shower, CORSIKA needs to know the primary particle type and energy, and its direction of propagation. These quantities are generally fed into CORSIKA by means of a customised driver script written in `perl`, which allows the air shower program to be run in batches of arbitrary size (limited only by the amount of available disk space). By calls to two helper programs, the script also permits air showers to be generated according to a given energy spectrum, passed to the script in the form of a spectral index α and an upper and lower energy bound. Furthermore, it is possible to generate ("throw") sets of CORSIKA showers with their actual directions randomised uniformly in solid angle around a central direction (*solid-angle scattering*). In this case, the driver script is given the desired mean heading of the primary particles and the half-angle of the cone within which the actual directions are to be scattered. The latter two features are convenient for simulating hadronic background, as cosmic rays arrive isotropically and with a (more or less precisely) known energy spectrum.

Apart from the main shower parameters, CORSIKA requires a large number of other input variables, such as the height above sea level of the observatory, the vector of the geomagnetic field at the observing site, the atmospheric model, the options for the various interaction codes *etc.* These are hardcoded in a template input file, which is used by the driver script along with the shower parameters to create a complete input card. A typical example of a CORSIKA input file is printed in Appendix A.1.

STACEE output format

Once all of an air shower's Cherenkov photons that survived the global efficiency cuts have been traced to the *observation level* (selected to be slightly higher than the altitude of the detector, and in STACEE's case 1740 m ASL), they are written to disk. CORSIKA writes a separate file for each simulated air shower, which can be rather large, in particular for air showers from gamma rays above a few 100 GeV. A special output

[5] Since the global detector efficiency cuts are already applied in the air shower simulation, CORSIKA shower output files generated with this cut cannot be used for studies where unbiased photon densities are required, *i.e.* without taking into account the optical detector components. In that case the efficiency values in the file "`quanteff.dat`" need to be set to 1.00 for all wavelengths.

routine for CORSIKA has been developed by the STACEE collaboration, which uses a very compact output format that keeps storage space requirements at a manageable level. The original format has undergone some minor changes over the course of its lifetime, and will now be described briefly.

CORSIKA's STACEE output consists of a bank called *shower header*, which is written in the beginning of each air shower file and contains important shower-specific information such as the primary particle type, its energy and direction, and a table with the efficiency cut values that were applied in CORSIKA. It also has several fields reserved for quantities that may be filled by subsequent helper programs, which we will not describe here. Following the shower header and making up the bulk of the output file are entries for all Cherenkov photons surviving to observation level. Each photon entry is made up of the following quantities:

- x and y landing positions on the ground (at the observation level), in STACEE coordinates[6] relative to the primary landing position (*i.e.* the shower core).
- n_x and n_y components of the unit wave vector of the photon, in STACEE coordinates, determining the *direction* of the photon.
- Arrival time t of the photon at the observation level. The absolute time has little significance (it is measured with respect to the mean arrival time of all the shower's Cherenkov photons, which may fluctuate significantly), only the *relative* time carries the physical information about the shower profile.
- The photon's wavelength λ.

The precision and dynamical range of these variables are limited by the exact photon packing scheme employed in the STACEE output routine. The standard photon output scheme is given in Table 5.3, and the associated ranges and resolutions are listed in Table 5.4. The compressed CORSIKA output can then be read in by custom programs such as STACEE's optical simulation, `Sandfield`.

5.3 Optical simulation: *Sandfield*

The STACEE optics simulation `Sandfield` (short for *SANDia FIELD simulator*) is responsible for simulating the transport of Cherenkov light from a horizontal, planar

[6]Here the coordinate axes are chosen so that the x-direction points to geographic East, y to geographic North and z vertically upwards, corresponding to the definition in all other STACEE software. Note that the origin is the shower core, whose position in relation to the "STACEE origin", which is the centre of the tower, is later randomised.

```
15 14 13 12 11 10  9  8  7  6  5  4  3  2  1  0
-----------------------------------------------
<------------- xpos (lower 16 bits) ---------->
<------------- ypos (lower 16 bits) ---------->
<------------- time (lower 16 bits) ---------->
<------------- nx (16 bits) ------------------>
<------------- ny (16 bits) ------------------>
th tg xh xg yh yg <----- lambda (10 bits) ---->

ref:   xpos = (xh xg xf xe xd .... x3 x2 x1 x0)
ref:   ypos = (yh yg yf ye yd .... y3 y2 y1 y0)
ref:   time = (th tg tf te td .... t3 t2 t1 t0)
```

Table 5.3: Scheme of the customised compressed CORSIKA Cherenkov photon output format used by STACEE. `lambda` refers to the wavelength of the photon, `(xpos, ypos, zpos)` is the position and `(nx, ny, nz)` the normalised momentum vector.

Quantities	Symbols	Dynamical range	Resolution
position	x, y	[−1310.72 m, 1310.71 m]	0.01 m
direction	n_x, n_y	[−1.0, 1.0]	3.05e−5
arrival time	t	[−4000 ns, 9107.15 ns]	0.05 ns
wavelength	λ	[0 nm, 1023 nm]	1.0 nm

Table 5.4: Dynamical ranges and resolutions of the CORSIKA photon bank quantities for the standard photon packing scheme (`CABE`).

surface above the heliostat field to the DTIRCs in the STACEE cameras. As such, it is practically a ray-trace program that reads CORSIKA output and processes it. While each CORSIKA file corresponds to the Cherenkov photons from only one shower, the number of CORSIKA showers read by `Sandfield` during one execution is only limited by the operating system specific file size limit for its output. Since only a very small fraction of the Cherenkov photons read from the input files end up releasing a (virtual) photoelectron from one of the PMT photocathodes, one `Sandfield` output file may contain the photoelectron information from hundreds to hundreds of thousands of CORSIKA showers, depending on the primary energy and type (and the shower core location). The output of `Sandfield` contains a list of photoelectrons for each PMT, along with their exact relative arrival times.

Processing an individual CORSIKA shower file begins with reading the shower header (see Sec. 5.2.4) and passing it on to the output file. Before starting the ray tracing for each shower, a shower core position needs to be chosen. The parameter input card

for `Sandfield` contains a variable called *scattering radius*, by means of which the user can specify the radius of the circle (or, for inclined air showers, the minor axis of the resulting ellipse) around the centre[7] of the heliostat array within which the shower core is to be chosen. The position is determined at random for each new shower, with uniform probability in area. The scattering radius must be chosen sufficiently large so that edge effects resulting from ignoring showers with large impact parameters that may still trigger the detector are avoided, yet not excessively large in order to keep the processing times at a manageable level.

Once a shower core position has been assigned, the photon entries are read sequentially, decoded, and translated laterally by the vector to the shower core. From the starting position, direction, and wavelength of each photon, `Sandfield` then calculates where the photon's trajectory intersects the ground. If this is outside of the heliostat array, the photon is immediately discarded; if not, the program checks if the reflective area of a heliostat is hit. Each heliostat is modelled in detail, *i.e.* as a grid of 5×5 square facets with interspersed gaps. The facets are perfectly parabolic and have focal lengths as determined by the geometry of the laser alignment scheme (Sec. 4.2.2).[8] They need not be perfectly aligned, however; the exact degree of facet misalignment is set via the `Sandfield` input parameter file. From comparisons of simulated light distributions with sunspots and laser images during facet alignment it was found that the random facet misalignment width should be $\sigma_f = 0.2\,\mathrm{m}$ for the 260-level and $\sigma_f = 0.1\,\mathrm{m}$ for the 220-level heliostats (Scalzo, 2000), measured at the position of the laser alignment target on the face of the tower.

Once a photon hits one of the heliostat facets, the reflected vector is calculated. `Sandfield` next checks to see whether the photon hits the PMT camera before reaching the secondary mirror, because despite the off-axis configuration, a substantial fraction of the secondary is obscured when viewed from the closer heliostats. The secondary mirrors are modeled as circular, perfectly spherical surfaces.[9] The simulated PMTs are aligned in such a way that the symmetry axes of the DTIRCs point directly at the centre of the mirror, with the front faces of the DTIRCs aligned at the mirror's focal length.

The vast majority of Cherenkov photons misses heliostats and secondary mirrors, and many of those that reflect off the secondary miss all DTIRCs. Only a small fraction

[7] The array centre is defined as the centre of gravity of all 64 STACEE heliostats, which is $(x, y, z) = (0.005, 117.092, 4.34)$ m. Here the x-axis points East, the y-axis North, and (0,0,0) is the centre of the base of the Solar Tower.

[8] In reality, the heliostat facets are not perfect paraboloids. See Sec. 5.5.2.

[9] Even though this is not strictly true for the secondaries on the 260-level (Sec. 4.1.2), the irregular outer edges of the mirrors are of little importance, as they are outside the PMT cans' fields of view.

of the original photons survive to the front face of the latter, which is where the ray-trace ends. Appendix A.2 contains an excerpt of a typical `Sandfield` input parameter file.

5.3.1 Enhancement look-up table

As mentioned in Sec. 4.1.3, DTIRC transmissivity as well as quantum efficiency of the PMT photocathode depend strongly on the angle of incidence of the photon, giving rise to an enhancement of the sensitivity at larger angles of incidence. The response of all cans (*i.e.* DTIRC-PMT combinations) to incoming light was measured in the laboratory as a function of angle of incidence and wavelength. In `Sandfield`, the values of (enhancement *times* quantum efficiency), divided by the global efficiency already taken out in CORSIKA, are tabulated as a function of DTIRC opening angle, photon wavelength and entrance angle into the DTIRC. Interpolations between the values from this look-up table are then used to decide whether an incoming photon releases a photoelectron or not. Note that this method also takes into account the angular dependence of the quantum efficiency.

The reflectivities of all secondary mirrors are assumed to be 90% in CORSIKA; intrinsic differences, obtained via reflectometer measurements at regular intervals, are taken account of downstream in the electronics simulation.[10] To process CORSIKA showers generated with the quantum efficiency cut, a corresponding option must be turned on in the `Sandfield` parameter file.

5.4 Electronics simulation: *Elec*

As the last stage of the simulation chain, the purpose of the electronics simulation (`Elec`) is the conversion of `Sandfield` output, consisting of nothing but a number of photoelectrons and associated times for each STACEE channel, to output quantities that resemble those of real data as closely as possible. Its main aim is thus the creation of "fake data" in the form of the same types of `stoff` banks that are filled during the acquisition of real data. The program furthermore ensures that all simulation input parameters are added to the output data stream in this same format. Finally, in order to allow for a calculation of effective areas, some details on those simulated showers that did *not* generate a trigger are provided in a separate output stream.

[10] This is somewhat counter-intuitive, as the reflectivity parameter clearly falls into the domain of the optics simulation. It has the potential advantage that the same Sandfield output files, which take much longer to be produced, could be used even after a change in the reflectivities. In practise the latter did rarely occur.

5.4.1 Operating principle

The fundamental algorithms on which `Elec` operates shall be described in the following. Upon opening a `Sandfield` output file, the program forwards to the first shower entry and reads the shower header bank. Selected quantities from the bank are stored in memory until a decision is made whether the event will be written out or not. The program then reads the subsequent list of photoelectron (PE) times for each PMT, and subtracts from these the channel-specific TOF delays appropriate for the simulated detector heading; these delays are calculated in the same way as it is done by the DAQ program (`vmedaq`, see Sec. 4.4.6). `Elec` then performs a preliminary inspection of the "shower" (from now on called *event*, since it has been reduced to a set of voltage signals), in order to determine if there is a remote possibility of it generating a trigger. This screening process is done in two levels of increasing event size (and thus likelihood of triggering). An overwhelming majority of events are rejected in this manner before entering the actual trigger simulation, saving precious computation time. For events which pass the inspection, an appropriate number of night-sky background (NSB) photoelectrons, determined by a set of 64 input PMT anode currents, are added to the *signal* PEs (*i.e.* the PEs due to Cherenkov light). The combined signal-plus-background list of PE times is then converted to a waveform consisting of linearly superimposed single photoelectron pulses. The single-PE pulses are constructed via an empirical pulse function obtained with low intensity laser measurements. Below in Sec. 5.4.2 we will discuss the utilised PE-pulse function in detail.

The mean pulse amplitudes of the single-PE pulses which make up a trace are determined by the selected gain G of each channel, under the requirement that the integrated area under the pulses (*i.e.* the charge) is $C = G\,e$. Amplitudes of individual PE-pulses are then randomised around the nominal pulse height according to a Gaussian distribution of width 30%, in order to take into account random gain fluctuations in the PMTs. Before being added to the trace, the times of those scaled single-PE pulse functions that are associated with signal photons are subjected to a Gaussian timing jitter of selectable width (by default 0.4 ns), with the purpose of modelling the PMT transit time spread. Optionally, a variable amount of modulated RF noise can also be added to the traces. The latter consists of a sinusoidal oscillation of frequency 96 MHz, modulated by a sine wave of 4 MHz, both random in phase, as determined empirically via oscilloscope measurements on a large number of channels.[11]

[11] This feature was included in the electronics simulation because some STACEE channels were affected by large-amplitude RF noise. The long signal cables from the PMTs to the filter boxes, even though

Via a discriminator model, which produces an output whenever a rising edge of the PMT waveform exceeds a given threshold, the final 64 waveforms are subsequently converted to a list of discriminator edges for each channel. Through a detailed simulation of the STACEE digital delay and trigger boards (MADDOG; see Sec. 4.4.4), a two-fold coincidence logic is then applied to the discriminator times, and a trigger decision is made. For all triggered events, the internal timing and waveform data are converted to the format required by `stoff`, and passed on to `stoff` subroutines for encoding them and writing them to disk. A designated Monte-Carlo `stoff` bank is also written, filled with quantities from the event header.

The electronics simulation may also be run in other modes, which do not require any air-shower related type of input; in fact all of `vmedaq`'s trigger modes are supported. Simulated laser runs may be produced, triggered via photodiode or by MADDOG, for instance in order to investigate the dependence of various gain calculation methods on timing and pulse shape. `Elec` may also run in free-running noise mode, where only triggers from NSB are counted, *i.e.* no signal pulses are added. This corresponds to the detector being run with scrambled (randomised) delays, and is essential for determining the breakpoint, *i.e.* the thresholds above which triggers from random coincidences are negligible.

5.4.2 Photoelectron pulse function

The exact form of the employed single-photoelectron pulse function is of some importance, as properties like the width-to-height ratio sensitively affect the likelihood that a given amount of deposited charge will exceed the discriminator threshold. For `Elec`, using an empirical parameterization of a PE-pulse was chosen over a library of measured pulses, mainly because recording a single-PE pulse in noise-free conditions proved to be rather difficult, but also because a functional form was more versatile and easier to adapt to actual changes in pulse shape, which may have happened as a result of changes in the electronics or signal cable lengths.

`Elec` uses one of two possible parameterizations of the single-PE pulse, which are:

$$F_0(t) = A_0 \frac{k_0}{\alpha} \left(\frac{k_0}{\alpha}\, t\right)^{1.5} e^{-(0.1\,(k_0/\alpha)\,t)^2} \tag{5.3}$$

$$F_1(t) = A_1 \frac{k_1}{\alpha} \left(\frac{k_1}{\alpha}\, t\right)^{2.37} e^{-1.32\,(k_1/\alpha)\,t} \tag{5.4}$$

shielded, were very susceptible to RF pickup. The exact origin of the RF is unknown, but it may have originated from a transmitter in one of the other installations at *Sandia National Labs*.

Here the A_i are normalisation constants, the k_i are dimensionless constants relating to the intrinsic integral-to-peak ratios of the functions, t is time, and α is the *pulse-shape parameter*. The latter is important, as it determines the integral-to-peak ratio of the pulse *independent* of the actual form of parameterization. The pulse-shape parameter is dialed in via the input parameter file, but is usually left constant at 4.0 ns, a value that was obtained through FGIT measurements (Sec. 4.5.2). The two `Elec` pulse functions are plotted in Fig. 5.3 for identical values of α. The choice of these two particular functional forms, F_0 and F_1, was inspired by works of Hinton (2001) and Schlenker *et al.* (2001), respectively.

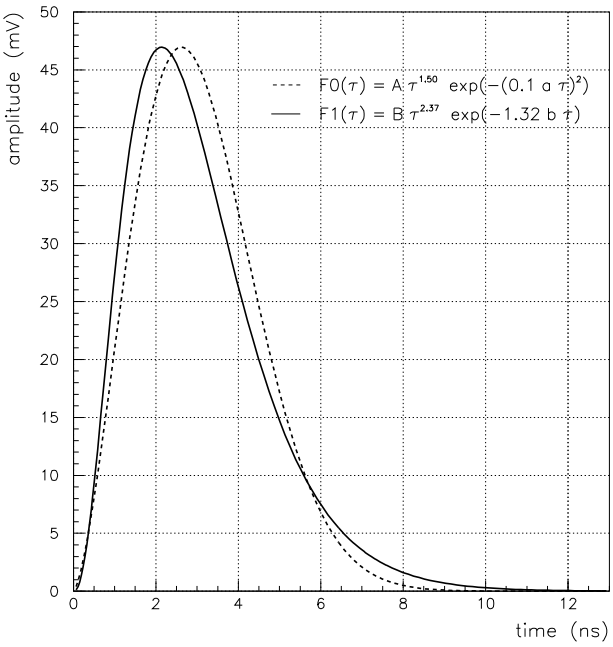

Figure 5.3: The two single photoelectron pulse functions available in `Elec`, normalised to have the same pulse height and integral (*i.e.* the same value of the pulse-shape parameter α).

A real single-PE pulse measured in the laboratory is shown in Fig. 5.4, after being passed through increasingly more modules of the STACEE electronics (data provided by Zweerink, 2004). It shows clearly that both amplifiers and filter box widen the pulse somewhat, and add a narrow overshoot. An oscillation (*ringing*) appears to set in right after the pulse independent of the traversed electronics modules, and is most likely due to the AC coupling. This effect also shows up in the autocorrelation functions of real FADC traces (Scalzo, 2004). Due to the low amplitude of the oscillation, it is probably not much more than a cosmetic effect. Along with the fact that the magnitude of the ringing changed between individual channels and its exact cause is not well understood,

these are the reasons why no attempt was made to model it within the simulation. The pulse width after all three modules, which is similar to the setup in the full experiment, is consistent with a pulse-shape parameter α of about 4.0 ns (*i.e.*, the default width in `Elec`).

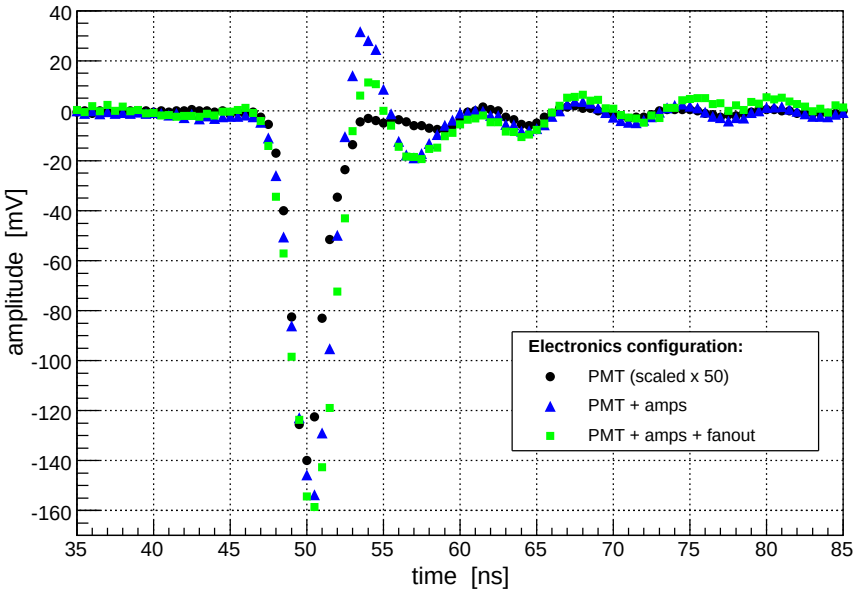

Figure 5.4: FADC traces of single photoelectron pulses recorded in the laboratory. *Black circles*: Raw pulse directly from the PMT, multiplied by a factor of 50; *blue triangles*: PE pulse after passing through UCD amplifiers, used by STACEE instead of the Phillips amps from 2005 onwards; *green squares*: PE pulse after passing through amplifiers and STACEE filter box. Data provided by Zweerink (2004).

Both pulse functions (F_0 and F_1) were found to yield identical trigger rates, provided the same value of α is used in both cases. With the somewhat crude time and voltage resolution of the STACEE FADCs during normal operation, and with the common high background currents, the two functions do not give rise to noticeably different shapes and RMS of the summed traces. $F_1(t)$ is slightly more asymmetric, and was used for the effective area simulations in this thesis.

The observation that in real FADC traces, many large pulses appeared to have long, slowly (several tens of ns) decaying tails, while simulated pulses decay more quickly, led temporarily to speculations that the particular shape of `Elec`'s single-PE pulse functions may not be appropriate, and that a more asymmetric pulse shape should be used. This explanation is highly unlikely, however, as measurements of single-PE pulses have consistently shown the pulses to be narrow and rather symmetric, much like the ones

shown in Fig. 5.4. It is of some significance that this disagreement seems to affect only large pulses, which are the sums of tens of photoelectrons. While the cause of this "pulse stretching" phenomenon has remained largely unknown, likely explanations include the AC-coupling downstream of the amplifiers, saturation effects in the electronics, or imperfections in the hadronic shower models. As an interesting alternative explanation, it may be that the known imperfections in the heliostat optics, such as warped and misaligned facets, and the resulting widening of the effective fields of view of individual heliostats allow the latter to view a larger portion of spread out Cherenkov showers from energetic primaries, giving rise to wider Cherenkov pulse timing profiles than predicted by `Sandfield` simulations. In that case, a better match for these energetic air showers could be obtained by increasing the facet misalignment parameters σ_f by a significant amount.

Input parameters

The electronics simulation is undoubtedly the module of the simulation chain with the most variable input; one may think of the necessity to simulate individual runs (or groups of runs), each requiring their own set of PMT currents and thresholds, or simulation studies to investigate the effect of changing certain run parameters on the performance of the detector. The main `Elec` input parameters, including all those that need to be changed in order to adapt a run to the given conditions, are read from a customisable input card. Basically all parameters entered into the real DAQ have their counterparts in `Elec`: trigger coincidence mode and multiplicity conditions, PMT thresholds and canting parameter. The source direction is usually read from the header bank in the `Sandfield` output file, but can be entered manually. A much larger number of input parameters are inherent to the detector itself, some of which are subject to constant change and are thus usually taken from a particular data run that is to be simulated; examples are the PMT currents and gains, and the discriminator thresholds. Another class of parameters are set to values that had been measured directly or indirectly, but were not expected to change much over time, such as the discriminator and TDC deadtimes, and the single-PE pulse width. Appendix A.3 contains a representative `Elec` input parameter file, along with explanations of all quantities. A set of constants, *i.e.* parameters which are not subject to change at all during normal running, are read in from a separate file.

Program output

The main `Elec` output is the "fake data" file consisting of `stoff` banks. These include both event and non-event banks, and basically all data fields are filled with appropriate

values that are either taken from the input card, from one of a number of constants files, or are calculated from the simulated data themselves. Apart from the `stoff` format output file containing the fake data, `Elec` produces a text output file with some basic information on all showers that are read from the `Sandfield` file, such as shower energy, direction, core position, number of created PEs and, most importantly, whether the shower generated a trigger or not. This file can be used to calculate trigger probabilities, and hence effective areas. The program also generates a useful text file with calculated PMT rates for each channel, along with the value of the input current.

5.5 Simulation cross-checks

Given the large number of input parameters at basically every stage of the simulation chain, some of which are difficult to measure, the simulations need to be "calibrated" to a certain degree. Naturally, this does not mean the introduction of *ad hoc* correction factors, which would be difficult or impossible to justify. However, our knowledge of many of the crucial input parameters is affected by considerable errors, resulting, for instance, from

- imperfections in the measurement technique
- performing our measurements only on a subset of a larger number of relevant devices (channels, PMTs, heliostats *etc.*)
- taking measurements only in a subset of many possible conditions, or configurations of the apparatus (*e.g.* detector heading, weather conditions *etc.*)

A number of simulation input parameters are in fact known with sufficiently large uncertainty, and these may be fine-tuned in order for the simulation results to match observable quantities. The most important examples for parameters that have a certain "degree of freedom" are the atmospheric model, the heliostat facet defocusing parameter, the exact discriminator latency (time over threshold requirement) and the exact PMT gains.[12]

Once the simulations are tuned well enough, the predictions of the Monte-Carlo may be compared in detail to measured quantities. This step is in fact required for validating the detector sensitivity forecast by the simulations, upon which STACEE's final fluxes and flux limits depend rather sensitively. It also serves the purpose of providing us with

[12]Note that most of these parameters are subject to measurement, but in reality these measurements were neither very accurate nor done very frequently.

an estimate of the systematic errors incurred within the detector Monte-Carlo. This section addresses some of the more important simulation cross-checks.

5.5.1 Air shower simulation cross-checks

Fortunately, the air shower simulation CORSIKA is used by a large number of collaborations working in experimental high-energy astrophysics worldwide, and hadronic interaction codes such as QGSJET, SIBYLL and FLUKA also by major high-energy physics experiments. The EGS4 package is even more robust, in particular since the physics of electromagnetic shower development are known to high precision. One may thus assume that any major inconsistencies in the program as a whole would have been discovered by now, even for the relatively low energy range of interest to STACEE. Nevertheless, some comparisons with other simulations have been made (Oser, 2000; Hinton, 2000), revealing minor differences which are consistent with different assumptions in regard to the atmospheric model. The error due to the latter alone gives rise to a 10% fluctuation of STACEE's predicted trigger rate (Sec. 5.2.3), and hence dominates the total uncertainty adherent to the air shower simulation part.

5.5.2 Sun and moon spots

The optical simulation is most certainly too idealistic, and thus probably too optimistic in regard to predicting the overall optical throughput of the heliostat–secondary mirror system. In the simulation, all reflective surfaces are perfectly smooth and uniform; in reality, the heliostat facets are weathered, sometimes cracked, and far from perfectly parabolic.[13] The simulation attempts to correct for this fact indirectly by assigning properly tuned heliostat reflectivities and facet misalignment parameters.

The easiest way to check the optical simulation is by comparing simulated intensity distributions to measured CCD images of the Sun or the Moon, both in terms of light distributions and absolute intensities. Images of the Moon on the camera plane, such as those shown in Fig. 4.16, can be compared to fake Moon images from `Sandfield` (Fortin, 2005). One way to compare absolute intensities of the optical system up to the heliostats (*i.e.* excluding the secondary optics) is by comparing real and simulated sunspot images (Sec. 4.2.1). Fig. 5.5 shows a comparison of measured radial intensity profiles with those output by `Sandfield`, under the assumption that $\sigma_f = 0\,\mathrm{m}$. Averaged over several simulated sunspots, the predicted percentage of light within a given radius

[13] In fact, a few of them have broken adjustment bolts and can no longer be focussed properly.

was found to agree with measurements to better that 5%, even with perfectly aligned simulated heliostat facets (see *e.g.* Oser, 1999).

Fig. 5.6 shows a more recent comparison of simulated and real sunspot data, under inclusion of randomised facet misalignments as described in Sec. 5.3, but without adding bias errors. The difference between actual and optimised efficiencies is that the former are measured within a circle centered on the fixed sunspot bias target, and thus include the effect of non-optimised biases. Despite most real heliostats being out of bias (upper left panel), simulated and measured efficiencies were found to match to within 5% (lower left panel). For any given heliostat, the results agree to within 11%, as we can see from the widths of the two fitted Gaussians.

5.5.3 Angular acceptance

Inconsistencies in the optical simulation as well as in the trigger model may show up as biases in the angular detector response function. The angular acceptance obtained from simulations should be more or less symmetric about the nominal detector heading, and the width of the distribution should be that of the detector's field of view. For the relatively large width of the coincidence gates imposed by the STACEE trigger, the detector field of view is determined by the optics alone. As we have seen in Sec. 4.1.3, the average field of view of individual heliostats is about 0.3° in half angle, when one assumes all heliostats to be point-like objects and that their optical components are perfectly aligned. Real heliostats however are extended objects, and together with the fact that the size of the Cherenkov light pool is smaller than the diameter of the heliostat field, the effective field of view of the detector as a whole can expected to be somewhat larger.

The angular response of the detector was studied using simulations of gamma-ray showers generated on a spectrum with index $\alpha = 3.5$, by varying the primary direction with respect to the detector heading in both elevation and azimuth direction. The angular distributions of the trigger probability are shown in Fig. 5.7. Two-sided *erf* function fits to both distributions suggest that the field of view as measured at FWHM is 0.8° in both directions, but the offset angle beyond which the trigger probability begins to fall off rapidly is somewhat smaller ($\sim 0.5°$). These values are in agreement with observations.

It comes as a mild surprise that while the azimuth scan shows a plateau within $\pm 0.5°$ from the pointing direction, as expected, the elevation scan reveals a significant enhancement near the edge of the field of view. This peculiar effect is caused by the

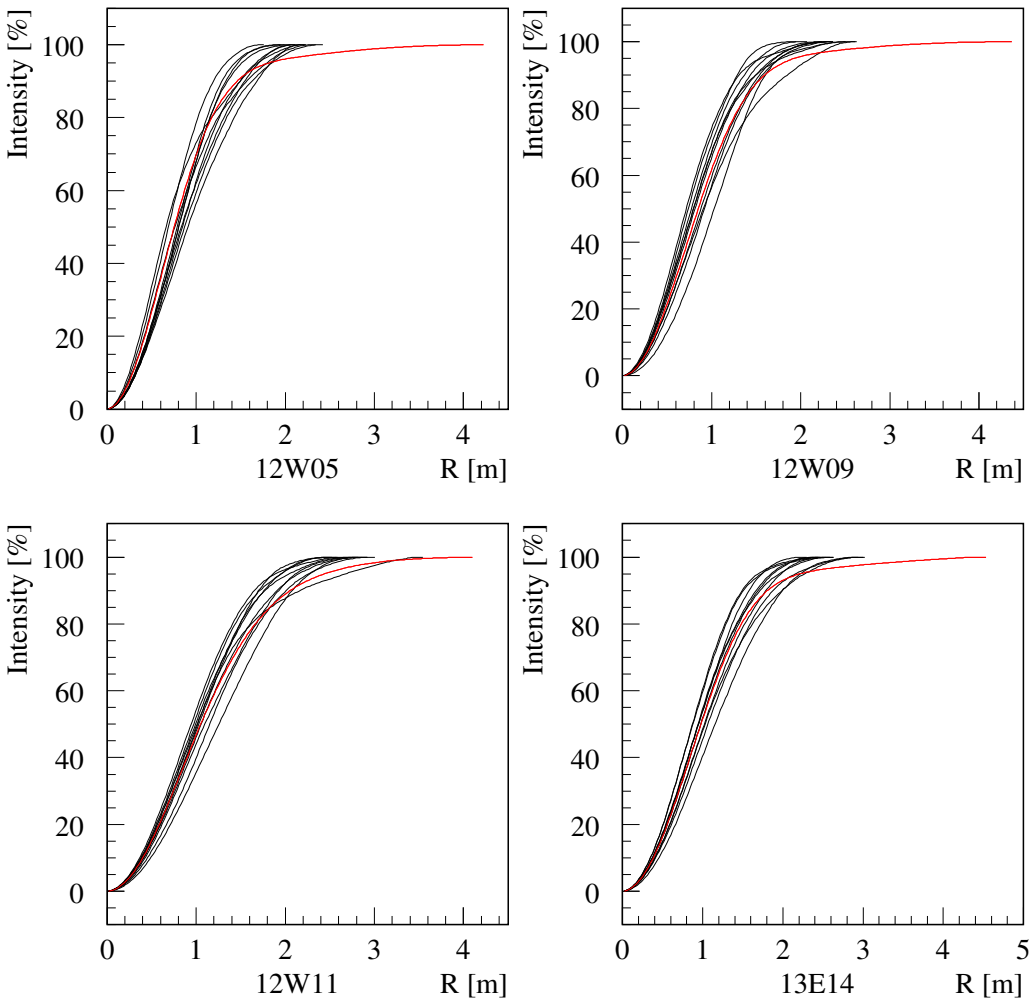

Figure 5.5: Comparison of measured sunspot intensity profiles (relative intensity *vs.* radius from the centre of the spot) to results from the optical simulation `Sandfield`. The single *red* curves are the CCD measurements, the set of *black* curves are results from different trials of the simulation. Shown are profiles for four different heliostats. Adapted from Oser (1999).

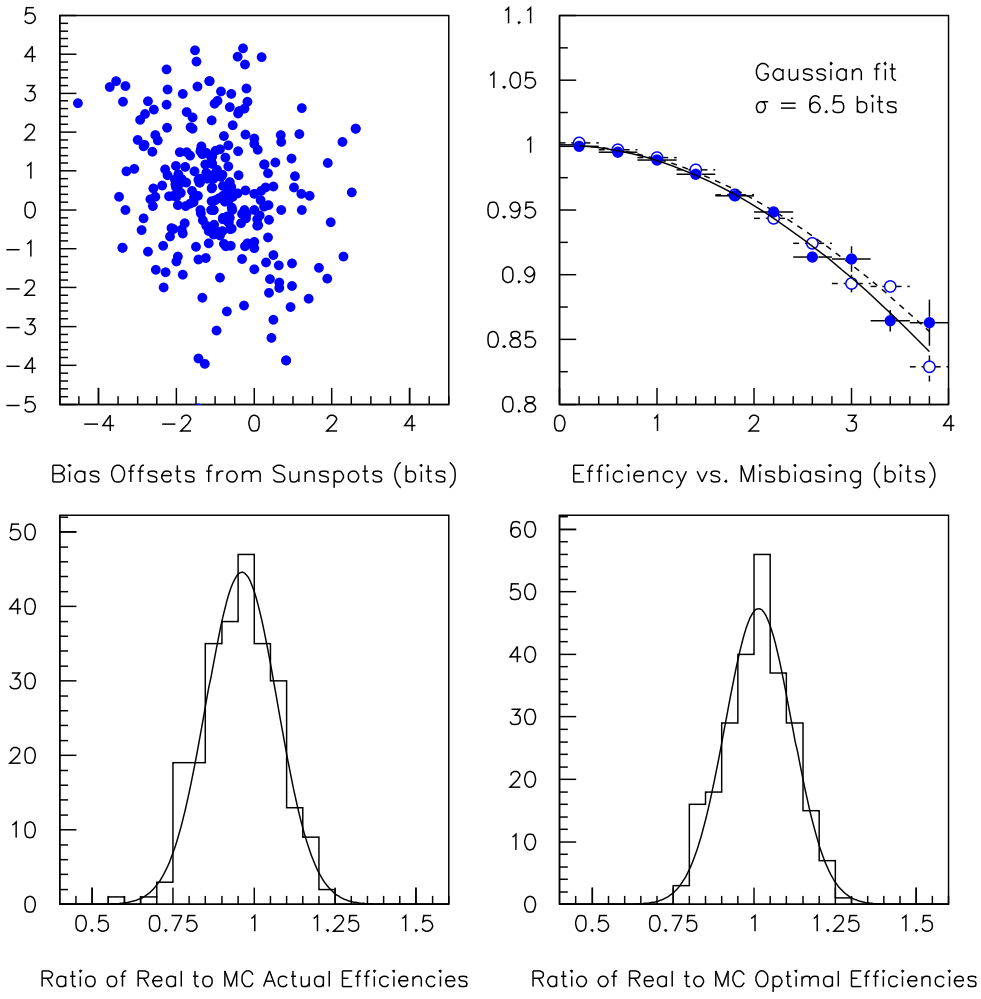

Figure 5.6: Comparison of real sunspot data to results of simulations from `Sandfield`. The sunspots were recorded between January and July 2003. *Top left*: Elevation *vs.* azimuth bias offsets for real heliostats inferred from sunspot centroids. *Top right*: Relative collection efficiency as a function of misbiasing from the commanded bias target, for real sunspots from CCD data (*solid circles, solid line*) and simulated sunspots (*empty circles, dashed line*). *Bottom left*: Ratio of real to simulated *actual* efficiencies (see text); the fitted Gaussian has a mean of $\mu = 0.96$ and a width of $\sigma = 0.11$. *Bottom right*: Ratio of real to simulated *optimised* efficiencies. The Gaussian fit has $\mu = 1.01$ and $\sigma = 0.11$. Taken from Scalzo (2004).

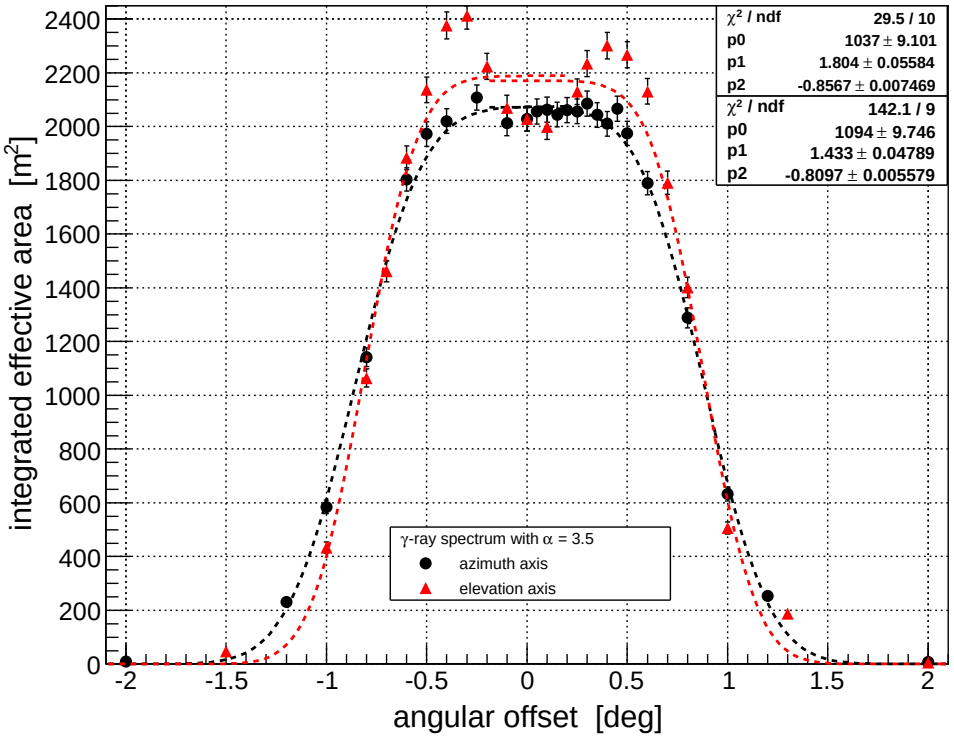

Figure 5.7: Simulated hardware pulse-spread function of STACEE, *i.e.* gamma-ray trigger probability as a function of angular offset of the primary direction from the detector heading. The source direction was taken to be the transit point of H 1426+428 (82.3° elevation, 0.0° azimuth), the spectral index was $\alpha = 3.5$, and heliostats were paracanted. The angular response was mapped in two directions, orthogonal in spherical coordinates: elevation (*red triangles*) and azimuth (*black circles*). Both distributions were fitted with two Error functions on either side of the peak, yielding good fits on the flanks and compatible widths of slightly more than 0.8° half-angle in both directions. Note that positive offsets on the elevation axis correspond to positive changes in elevation. Trigger probability is given in units of effective area integrated over the input gamma-ray spectrum.

enhancements of the DTIRC-PMT units (see Sec. 4.1.3). The reason why there is only a slight hint of this feature in the data for the azimuthal scan may be due to the different geometries involved; the cameras are grossly off-axis in elevation direction, but not in azimuth. Similarly, the slight asymmetry of the elevation curve can be explained by changes in the amount of camera box occultation, and the fact that for the chosen source position (north of zenith), the heliostat-secondary geometry becomes more off-axis for increasing zenith angles.

5.5.4 Discriminator rates

A reasonably robust test of the electronics simulation is the comparison of simulated to real discriminator rates for data where the PMT anode currents, effective discriminator thresholds and PMT gains are known to sufficient precision. The PMT rates are extremely sensitive to small changes in any of the aforementioned parameters (see *e.g.* Scalzo, 2004), as well as to the exact way the discriminators respond to PMT pulses that have amplitudes comparable to the threshold voltages (time over threshold requirement). While the PMT currents and thresholds are known reasonably well for most data runs, the gains have been known to fluctuate with time and may thus be affected by significant errors. Using a number of runs from different observing nights should reveal any scatter due to random gain offsets, albeit not due to systematic biases in the gain measurement procedure.

A comparison of measured and simulated PMT rates for four stable 28-minute runs can be seen in Fig. 5.8. The width of the band in the scatter plot is caused by the inherent uncertainties in the gain and discriminator threshold measurements, which are assumed to be about $0.1 \cdot 10^5$ and 2 - 4 mV, respectively, but probably also by systematic effects in the PMT current monitoring. While the centroid of the distribution is very well centered at a ratio of one for typical, *i.e.* intermediate and high PMT anode currents, the simulation consistently underpredicts the PMT rates for channels with very low currents. Given that predictions are on the mark for "normal" currents and the traces are always treated in exactly the same manner within the simulation (*i.e.* irrespective of current), the reason for the discrepancy must be sought in the detector electronics, which may have given rise to subtle effects that are not understood. Possibilities include extra pulses triggering the discriminators at constant rates, and which are not accounted for. Perhaps systematic effects in the current measurement play a role; while the precision of the measurement is estimated to be 1 - 2 μA, the exact response of the PMT anode circuit consisting of PMT base and Scan-ADCs to various true input currents was never measured.

Finally, significantly higher levels of RF noise than the low amplitudes dialed in for the simulation (1.5 mV RMS across all channels) could have caused the discriminators to fire at higher than expected rates, an effect which would also become more pronounced with decreasing PMT currents.[14] Those data points in Fig. 5.8 with significantly higher

[14] As mentioned previously, a highly variable (in time and across individual channels) level of RF interference was present in all channels. Due to the difficulty of keeping track of the actual amount of RF noise on each channel, as well as the transient nature of occasional large-amplitude "RF fits", these are generally not reflected in the choice of input parameters for the electronics simulation.

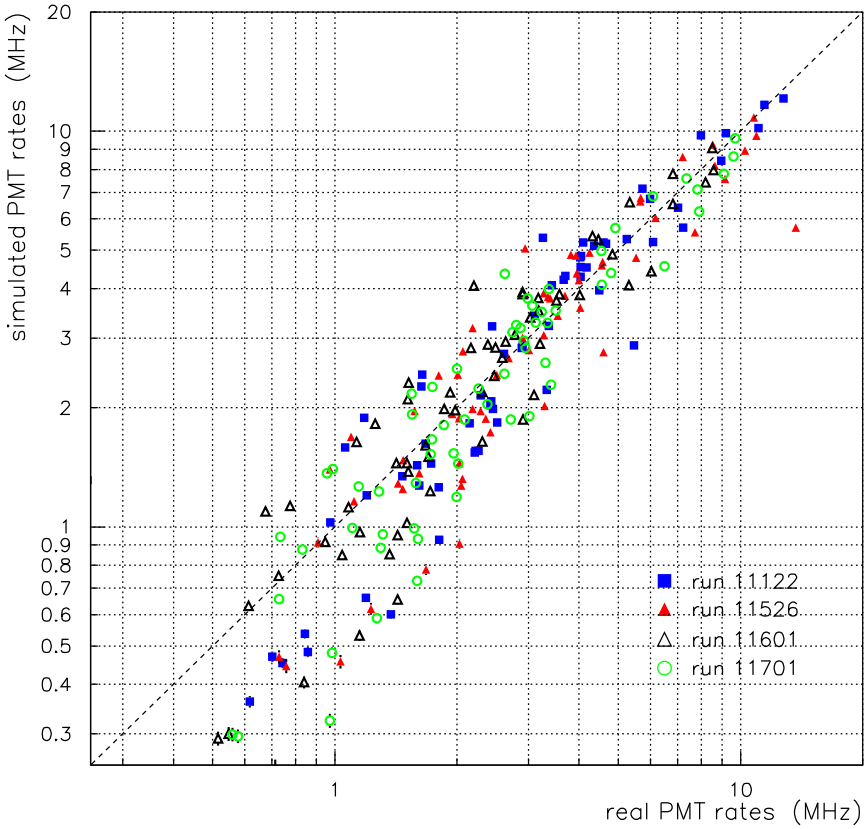

Figure 5.8: Scatter plot of simulated *vs.* measured PMT rates for four representative, stable 28' runs (*various symbols*, see legend). In an attempt to mimic each run, the electronics simulation was fed with the four sets of 64 measured anode currents, and channel-specific gains and discriminator thresholds as stored in the database. The required time-over-threshold for the discriminators was set to 0.8 ns. Statistical errors are too small to be visible. The centroid of the distribution is very well centered at a simulated-to-real rate ratio of unity (*dashed line*) for typical PMT currents, but for very low PMT currents the simulated rates fall short of the measured ones. Note that a few channels have significantly higher measured PMT rates than predicted; this is due to the presence of additional, large-RMS RF interference in the real data.

measured PMT rates are due to channels/runs that were affected by temporary large-amplitude RF interference.

5.5.5 *Cosmic-ray event trigger rate*

In the absence of a calibrated and sufficiently strong source of gamma rays, the overall response of air-Cherenkov telescopes may be checked simply by comparing the measured

cosmic-ray trigger rates to those predicted by detector simulations (see *e.g.* LeBohec & Holder, 2002). Naturally, this method does not provide the necessary accuracy for unambiguously determining the energy scale of the experiment, since the fluxes of the individual cosmic-ray nuclei in the energy range relevant for STACEE ($\sim 100\,\mathrm{GeV}$ – $10\,\mathrm{TeV}$ per nucleon) are not yet known to high enough precision. However, obtaining a match between simulated and measured cosmic-ray rates, ideally for several different detector headings, would lend credibility to all components of the simulation, including the particular choice of atmospheric model used.

To this aim, one must first decide on which cosmic-ray primaries to take into account, and find a spectral interpolation in the form of a power law for each of the particles to be simulated. Apart from the dominant protons and helium nuclei, we have here investigated the expected trigger rates due to the next most abundant cosmic-ray nuclei, the *CNO*-group, as well as cosmic electrons/positrons. The latter, although of very low flux above a few GeV, represent a potentially very problematic background; the resulting air showers are purely electromagnetic and thus indistinguishable from air showers initiated by gamma rays.

Cosmic-ray fluxes in the VHE energy range are usually measured by means of particle detectors onboard balloons, which ascend to altitudes between 35 and 40 km. The overall uncertainties in the integrated fluxes for each experiment are considerable (for the most recent BESS measurement, one of the most accurate so far, they are $\sim 15\%$ for protons and $\sim 20\%$ for helium; Haino *et al.*, 2004), and the scatter between the results of different experiments is even larger. The uncertainties are particularly large in the multi-TeV range, where only very few measurements exist. Fig. 5.9 shows cosmic proton and helium fluxes according to measurements up to 2002. The p and He primary spectra – normalisation and spectral index – used here are interpolations of the world data at VHE energies according to Hinton (2002), and are indicated by the dotted lines in Fig. 5.9. More recent results from BESS (Haino *et al.*, 2004) are compatible with this interpolation. The spectra of C and O were taken from ATIC (2005), and the relative normalisation of the N contribution from Gaisser & Stanev (2000). Parameters for the cosmic e^-/e^+ spectrum were obtained from an interpolation of the recent cosmic electron flux measurement by ATIC-2 (Chang *et al.*, 2005), scaled up by 10% in order to account for the positron flux, which is about 10% of the electron flux. The various spectral power-law interpolations are listed in Table 5.5.

In order to calculate the expected trigger rate ($\mathrm{d}N_i/\mathrm{d}t$) from a set of trigger probabilities P_i for each primary type i, where the P_i are simply the number of triggered events divided by the total number of generated showers (both provided by the `Elec` output),

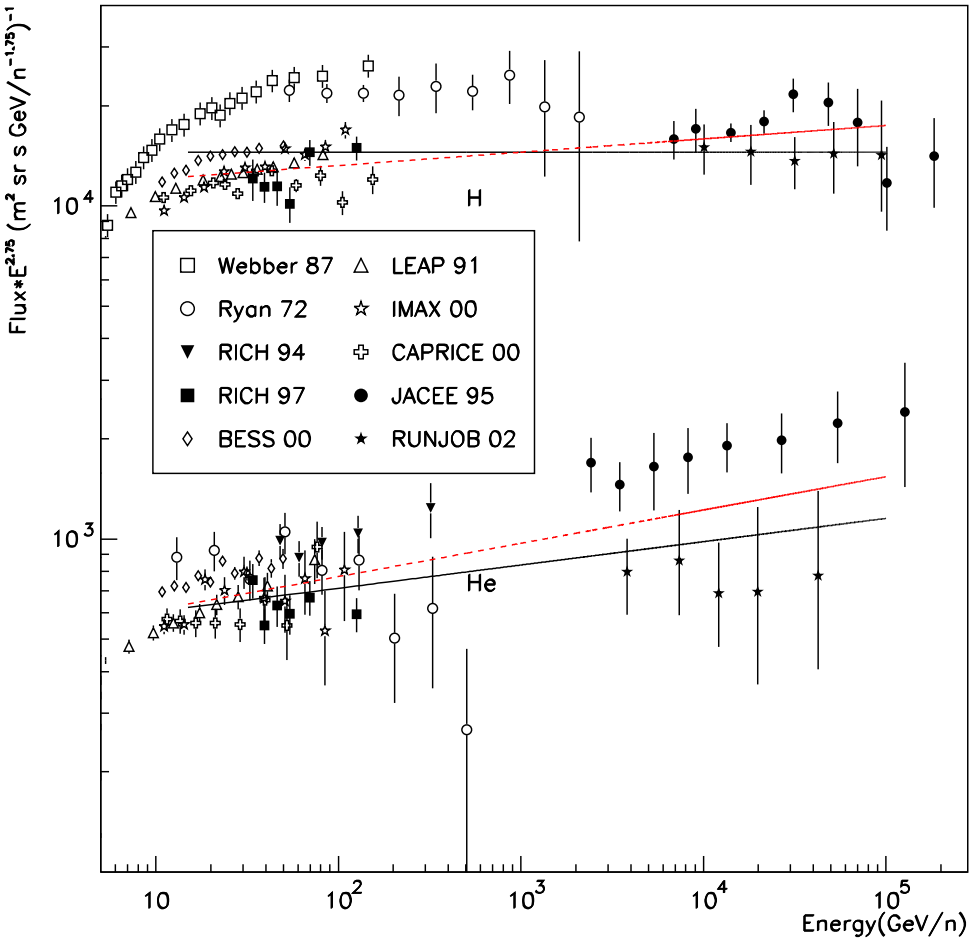

Figure 5.9: Measurements of the flux of cosmic protons (H) and helium nuclei (He) at energies relevant to STACEE by various balloon experiments. The black solid and red dashed lines are power-law fits to the data; the lines in red are the fits used as input to the CORSIKA air-shower simulation. Note that the energy unit is GeV *per nucleon*. From Hinton (2002).

$P_i = N_{i,trig}/N_{i,total}$, the latter must be multiplied by the geometrical acceptance factors η_i and the total integrated fluxes Φ_i above the energies $E_{0,i}$:

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = \eta_i \, \Phi_i \, P_i \tag{5.5}$$

with the definitions

$$\eta_i = \pi \, R^2_{\mathrm{scat},i} \cdot 2\pi \, (1 - \cos(\delta\theta_i))$$

$$\Phi_i \quad = \quad N_{\mathrm{int},i} \left(\frac{E_{0,i}}{1\,\mathrm{GeV}} \right)^{-(\alpha_i - 1)}$$

where the $R_{\mathrm{scat},i}$ are the core scattering radii and the $\delta\theta_i$ are the half-angles of the solid angle cones within which the shower cores and primary particle directions, respectively, were randomised uniformly. The $E_{0,i}$ are the given lower energy bounds of the generated spectra, which were chosen high enough so as to avoid excessively long processing times, and low enough so that the trigger probabilities are essentially zero at energies below them. The above expression for the integral fluxes Φ_i is just an integration of the simple power law (Eq. 2.31), with the two normalisation factors being related via

$$N_{\mathrm{int},i} = \frac{N_{0,i}}{(\alpha_i - 1)} \cdot \mathrm{GeV} \tag{5.6}$$

Since the interaction products of hadronic primaries – in particular of nuclei with higher Z – may possess substantial transverse momenta, hadronic showers occasionally give rise to pools of Cherenkov light relatively far from the shower core. Therefore, when simulating hadronic showers, sufficiently large scattering radii R_{scat} of 500 - 700 m and scattering solid-angles $\delta\theta$ of 4.0° - 4.5° were chosen for p/He and CNO primaries, respectively.[15] In terms of the total number of CORSIKA showers produced for this particular study, the numbers per primary particle are: $2.769 \cdot 10^6$ protons, 872 000 ^{4}He, 384 722 e^-, and 804 090 ^{12}C/^{14}N/^{16}O. Each CORSIKA shower was used 10 times with randomly chosen shower core (see Sec. 5.3), meaning that the figures quoted above need to be multiplied by 10 in order to obtain the actual numbers of simulated showers thrown.

The heliostat and electronics configuration used for the simulations was that of spring 2003, and the detector heading was taken to be the transit point of H 1426+428 (82.3° elevation, 0.0° azimuth). The resulting rates are shown in Table 5.5, along with the primary-specific simulation input parameters. The total predicted trigger rate is (5.238 ± 0.128) Hz, and is in good agreement with the average actual trigger rate given in Table 5.6.

In order to check if the simulations were able to reproduce the angular dependence of the L2-trigger rate as well, a second set of simulated cosmic-ray showers was generated for the detector heading corresponding to an hour angle of 30° of our source (H 1426+428), *i.e.* for an azimuth of 297.40° and an elevation of 65.54°. The parameters, in particular currents and thresholds, that were used as inputs for the electronics simulation were obtained by averaging over all live time of the H 1426+428 (2003) data set that falls

[15] When choosing cutoff values smaller than these, a significant fraction of the triggering showers are missed, in particular for the angular scattering.

Cosmic-ray spectra and simulated rates						
Primary particle	Φ_0 $[(\mathrm{m}^2\,\mathrm{s}\,\mathrm{sr}\,\mathrm{GeV})^{-1}]$	α	R_{scat} [m]	$\delta\theta$ [°]	E_0 [GeV]	$\mathrm{d}N/\mathrm{d}t$ [Hz]
p	11400	2.71	500	4.0	50	3.648 ± 0.114
He	5520	2.65	500	4.0	100	1.496 ± 0.059
CNO	2810	2.65	700	4.5	1000	0.043 ± 0.001
e^-/e^+	443	3.25	300	3.0	25	0.051 ± 0.002
SUMMED						5.238 ± 0.128

Table 5.5: Power-law interpolations of the VHE spectra of various cosmic-ray primary particles, simulation parameters used for each primary, and the predicted trigger rates for STACEE when observing the H 1426+428 transit point (with statistical errors). The simulated detector configuration corresponds to the observations of H 1426+428 in spring 2003 (monocanting). The interpolated spectra are of the form $\mathrm{d}N/\mathrm{d}E = N_0 E^{-\alpha}$.

into a given 10° hour-angle bin (see Sec. 7.3). The real L2-rates against which we want to compare the simulated rates are hence the average rates of all data corresponding to the same hour-angle bin. The results are summarised in Table 5.6, whereby the quoted errors reflect the scatter of individual live time intervals' L2-rates around the mean. Fig. 5.10 shows a plot of the averaged rates as a function of hour angle, along with the two data points predicted by the simulations. Considering the large error bars of the real L2-rates, resulting from changing discriminator thresholds, a certain drift of the PMT gains and varying atmospheric conditions, the agreement between simulated and real trigger rates is excellent.

Averaged L2 rates	
HA [°]	dN/dt [Hz]
0	5.155 ± 0.302
10	5.337 ± 0.719
20	5.123 ± 0.638
30	4.476 ± 0.259
40	4.777 ± 0.240
50	3.511 ± 0.184

Table 5.6: Averaged L2-rates of the H 1426+428 data set of 2003, binned according to hour angle (HA) in 10° steps. The statistical errors are standard deviations describing the scatter of the rates of individual live time intervals within each hour-angle bin.

It should be noted that relatively small changes in PMT gains, in particular when correlated across all channels like for instance due to the observed gain drifts over the course of an observing season, will translate to large changes of the L2 rates. In Fig. 5.11,

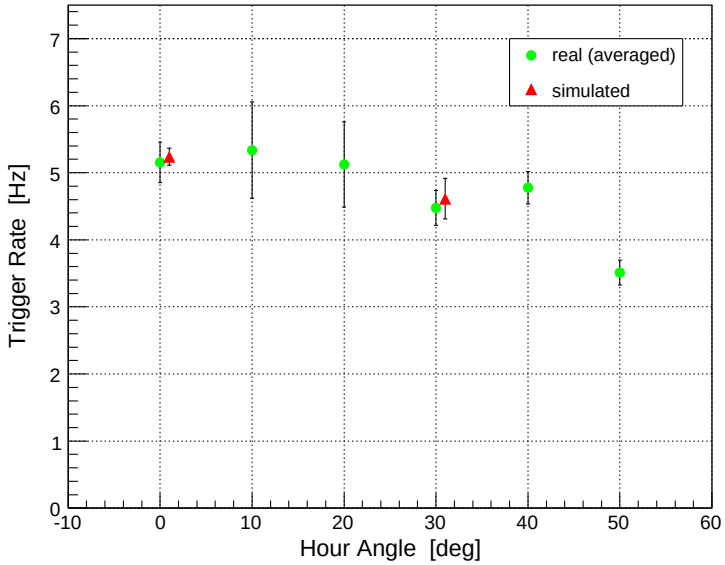

Figure 5.10: Averaged real (*green circles*) and simulated (*red triangles*) hadronic trigger rates as a function of hour angle for the 2003 data set of H 1426+428 ($RA = 217.14°$, $\delta = 42.67°$). The simulated rates are plotted offset by $+1°$ in HA for better visibility. Statistical errors are shown; for the real data they are standard deviations describing the scatter of individual live time intervals' L2-rates within each hour-angle bin.

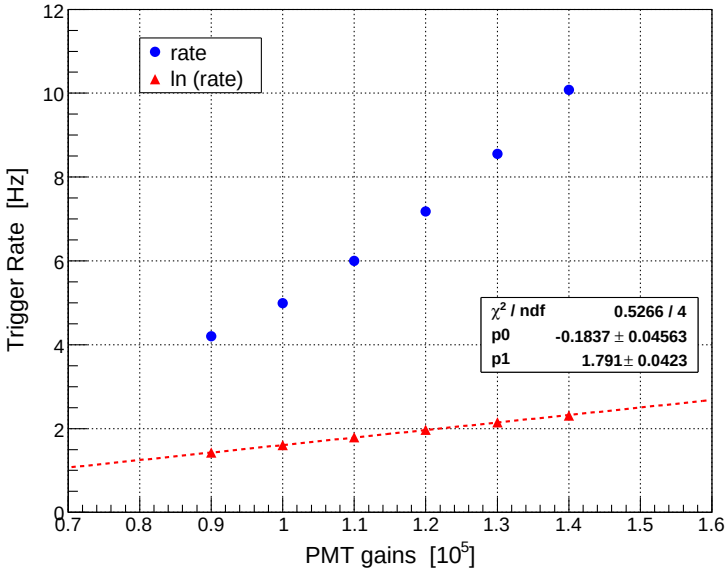

Figure 5.11: Simulated hadronic trigger rate as a function of global PMT gains. Statistical errors are plotted, but too small to be visible. Also shown are the logarithms of the rates, including a linear fit (*red triangles and dashed line*).

simulated L2 rates due to hadronic showers sampled by use of the standard spectrum are plotted as a function of the global PMT gains. As can be seen from the linear fit to the logarithms of the rates, the closeness to an exponential dependence is remarkable.

5.5.6 Canting curve

Another good end-to-end test of the simulation chain, which is sensitive to potential problems in air shower, optical and electronics simulations all at the same time, is the dependence of the STACEE hadronic trigger rate on the canting parameter. Fig. 5.12 shows overlaid real and simulated *canting curves*, consisting of trigger rates due to Cherenkov showers from cosmic rays with the detector pointed at zenith, under variation of the canting parameter. Both real and simulated data were obtained with all heliostats canted (monocanting). The simulated rates were obtained by using the appropriate admixtures of proton and helium showers with the standard spectra as given in Table 5.5. The match between real and simulated trigger rates is obviously quite good.

It should be noted that the PMT gains for the runs that make up this canting curve were known only approximately, as the data were taken before systematic monitoring of the gains began. In the simulation, the PMT gains were therefore adjusted until the absolute rates approximately matched those of the data. The different gains and thresholds, corresponding to a different (earlier) observing season, are also the reason why the trigger rates do not match the ones quoted in the previous section.

Looking at Fig. 5.12 more closely, we notice that for hadrons, the effective area is maximised at canting parameters larger than the nominal 0.08 km^{-1}, *i.e.* for the heliostats pointing at lower altitudes than for normal observations. Simulation studies for common gamma-ray spectra have shown that the rate due to *gamma rays* is indeed maximised at the nominal canting parameter.

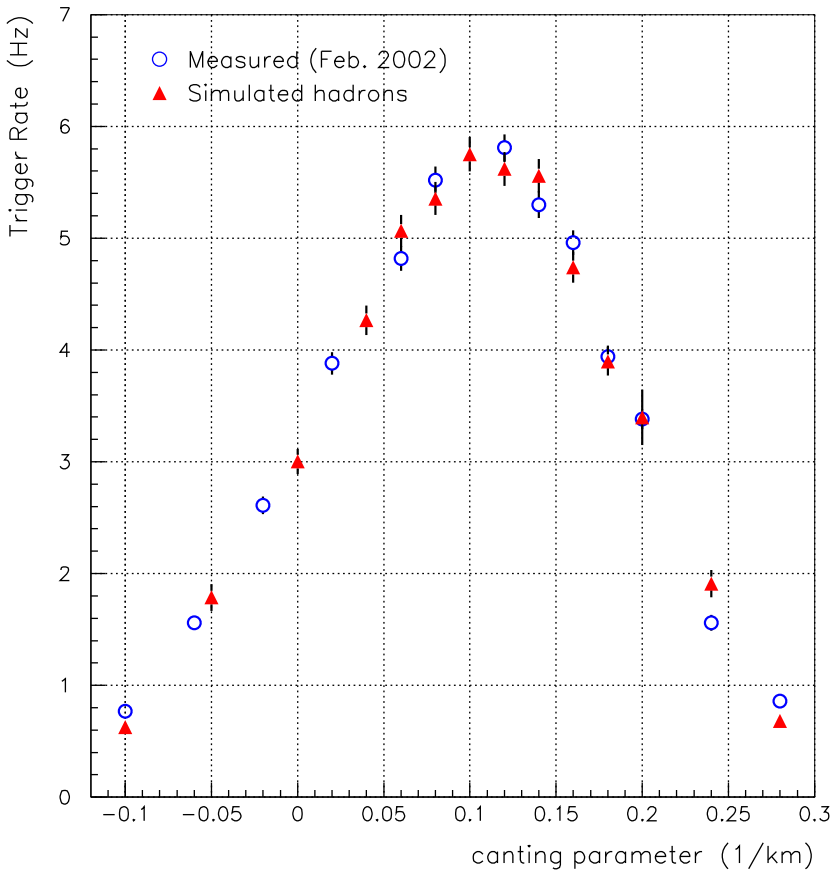

Figure 5.12: STACEE hadronic trigger rates at zenith as a function of the heliostat canting parameter (*canting curve*). Shown are a measured curve recorded in February 2002 (*empty blue circles*) and the L2 rates predicted by simulations of proton and helium nuclei showers using the standard spectra (see text; *red triangles*). Statistical errors are included for both data sets. Canting curves are for monocanted heliostats.

CHAPTER 6

DATA ANALYSIS

Being primarily – although not entirely – a counting experiment, analysis of STACEE data consists of two stages: extracting a clean sample of on-source and off-source data with known live times and identical background conditions, and examining the data event by event in order to reject hadron candidates and thus improve the signal-to-background ratio. In this chapter we will describe the individual steps that are necessary for a proper interpretation of the raw data, yielding a rate of gamma-ray candidate events.

6.1 The STACEE data

Before embarking on a description of the data analysis process itself, it is necessary to address the nature of the data, *i.e.* what quantities are recorded, and under which conditions. This will be done in the following two sections.

6.1.1 Observing technique

In order to understand the observation principle of Air Cherenkov telescopes (imaging ACT as well as wavefront-sampling detectors such as STACEE), and thus how the acquired data are to be understood, we shall begin with a description of the ON/OFF observing technique. Due to the overwhelming background of Cherenkov light flashes from cosmic-ray showers, observations of potential gamma-ray sources with ACTs require observations of a background region (which can be either simultaneous or alternating with the source observations), in such a way that the trigger rates of the latter can be subtracted from those of the corresponding *on-source* observations. By definition, the background region must not contain a significant source of gamma rays. For identical conditions and matching *live times* (the net times that the detector was active during each observation, *i.e.* total elapsed time minus deadtime due to data readout *etc.*), any difference in the number of events or in the trigger rate (*on-source excess*) would then have to be due to gamma rays from the prospective source.

For STACEE, this was done by performing separate on-source (henceforth *ON*) and off-source (*OFF*) observations. Each observation run lasted 28 minutes, so that a complete ON/OFF *pair* consisted of two 28-minute runs and two minutes in between runs, which were necessary to redirect the heliostats, change DAQ parameters *etc.*. In order for the ON/OFF method to be valid, hadronic background as well as night-sky background must be equal in both halves of a pair. Therefore, OFF observations were performed over the same local sky coordinates (in terms of azimuth and elevation) as ON observations. For a time difference of 30 minutes between the starts of the ON and OFF runs, the target coordinates for the OFF observation were thus offset 30 minutes in RA from the source coordinates, with identical declination. In order to maximise the observing time at higher elevations, where the night-sky background is lower, OFF runs were taken *before* the ON runs while the prospective source was rising in the sky, and *after* the ON runs when it was beyond transit and setting.

As a first step in data analysis, each run is analysed separately and some important quantities are reconstructed from the raw data (see Sec. 6.2). In a second step, both halves of a pair are examined together, and final quantities such as excess rate are determined (Sec. 6.4).

6.1.2 STACEE data format

The basis of all STACEE analysis software is a system of data packing and unpacking routines called `stoff` (short for *STACEE offline*). These routines store the data in binary format, organized in individual *banks* which contain data fields (variables) that are logically connected. The DAQ programs (`vmedaq` and the DAQs for the four FADC crates, see Sections 4.4.6 and 4.4.5) stored acquired data in the form of `stoff` objects, as does the last stage of the simulation chain (`Elec`) for creating simulated data. Non-event data that was acquired separately, such as heliostat status information, star photometry data and weather data, was also converted to `stoff` banks and merged with the mainstream data. All analysis programs are compiled with the `stoff` classes and are thus able to access these data files. Details about the original version of `stoff` can be found in Covault (2000), and an almost up-to-date list of the various banks in Scalzo (2004).

Each STACEE data file is essentially a sequential concatenation of individual banks, filled with values from the various electronics modules as well as with calibration data stored on disk. Note that the following description of data quantities makes sense only in combination with the description of the corresponding hardware, as presented in Chapter 4. For the purpose of this thesis, it suffices to point out that there are banks

which contain data applicable to a whole run (containing quantities such as run number, source number, trims, GPS time at start and end of the run, *etc.*), banks which occur once for each triggered DAQ event (*event banks*, details on which will follow below), and banks containing quantities that are measured in periodic intervals (*environment banks* and *non-event banks*) such as PMT currents, scaler values, TOF delay settings, heliostat status and weather information. Most banks were written by the main DAQ and the FADC DAQs over the course of the run; only the non-event banks were generated and merged with the data file later. Furthermore, `stoff` contains banks which are added and filled during offline reconstruction by the `Pass0` and `Pass1` analysis programs (see Sec. 6.2), and it provides support for the simulations by means of a bank reserved for simulation input parameters.

The event banks are the actual "heart" of the data, containing the information pertaining to the events that triggered the detector. These include:

- event GPS time
- type of event, *i.e.* where the trigger was generated: real (MADDOG), calibration (laser photodiode) or fake (pulse generator)
- values of all L1 (cluster rate) scalers
- raw timing information from vernier codes (MADDOG) and TDCs for all channels
- 192 ns of digitised FADC waveforms for all channels
- information relevant for the FADC calibration, including pedestal and RMS calculated from another 400 ns of FADC trace not written to disk

Other event banks are written by the analysis programs `Pass0` and `Pass1`; they are filled with reconstructed quantities, which will be described in Sections 6.2 and 6.6.

6.2 Basic reconstruction

The purpose of the first stage of data analysis is to convert the raw data, consisting of hardware (*i.e.* uncorrected) times, FADC counts, binary vernier codes (Sec. 4.4.4) and a number of other variables into physical quantities, such as absolute times and pulse charges. Also, a simple attempt is made to reconstruct the original Cherenkov shower by fitting spherical, parabolic and planar wave fronts to the hit times of individual channels. This is done by the program `Pass0`, a highly modular program written in C++. Classes defined in and used by this program are also used by the higher level analysis programs

such as `Pass1` and `stoff_pair`. Within `Pass0`, all calibration constants are taken into account, such that its output is (or should be) independent of the particular state of the detector at the time the data were taken. Another important analysis step that is done by `Pass0` is the equalisation of background light levels, which we shall describe in Sec. 6.3.

The most important tasks that are performed within `Pass0` on an event-by-event basis are:

1. The global timing offset resulting from the discretely stepped (at 32 Hz; see Sec. 4.4.4) trigger boards is found by averaging over all vernier times; this average is also called *vernier peak*. Fig. 6.1 shows typical vernier time distributions after correction for the vernier peak.

2. The FADC traces as well as the associated pedestal and RMS values are converted from *digital counts* to mV; FADC traces are corrected for the pedestal.

3. The effective PMT thresholds are calculated by discriminating the FADC waveforms in software at a variable threshold and comparing the software triggers to existing vernier hits (Sec. 4.5.4).

4. FADC traces are corrected for the differences in RMS between the ON and OFF runs of a pair (*padding*), in order to equalise the background light levels. The details of this process will be presented in Sec. 6.3.2.

5. The FADC waveforms are discriminated in software using the effective thresholds calculated earlier, or in cases where the latter could not be determined, the nominal thresholds. A *hit* is asserted for every falling-edge[1] threshold crossing. A slewing correction is applied to the hit times to account for different pulse heights.

6. The FADC times are corrected for channel-specific offsets (trims) as well as the event-by-event trigger jitter by subtracting the vernier peak found earlier. Then, using the corrected vernier times and the programmed delays, the expected position within the trace of a spherical Cherenkov wavefront originating at the heliostat canting point is calculated. This is referred to as the *dead-reckoned* position.

7. In a similar fashion, the raw hit times of the common-stop TDCs are converted to corrected times, and a dead-reckoned position is calculated. Since the TDCs are no longer used in the analysis, the details of this process will not be described here.

[1] Recall that the voltages of the PMT signals are negative.

8. Any FADC hits outside a so-called *in-time window* of 16 ns width, symmetrically centered on the dead-reckoned position, are discarded. For in-time hits, the corresponding heliostat hit times are calculated by subtracting the programmed delays and trims. To these *times on heliostat*, a wavefront parameterization (spherical, parabolic and planar) is fitted.

9. The FADC waveforms are integrated from 6 ns before to 10 ns after the dead-reckoned position of the wavefront on every channel, including those with no hits. The result is an estimate of the deposited charge.

10. The accumulated live time from run start to each event trigger is calculated from the deadtime clock and stored with the event data.

The `Pass0` wavefront fits are preliminary, as they do not yet include a fit to determine the shower core (*i.e.* the position where the extension of the primary particle direction intersects the ground), without which the shower direction cannot be unambiguously determined. More sophisticated fits will be described in Sec. 6.6 as part of advanced reconstruction.

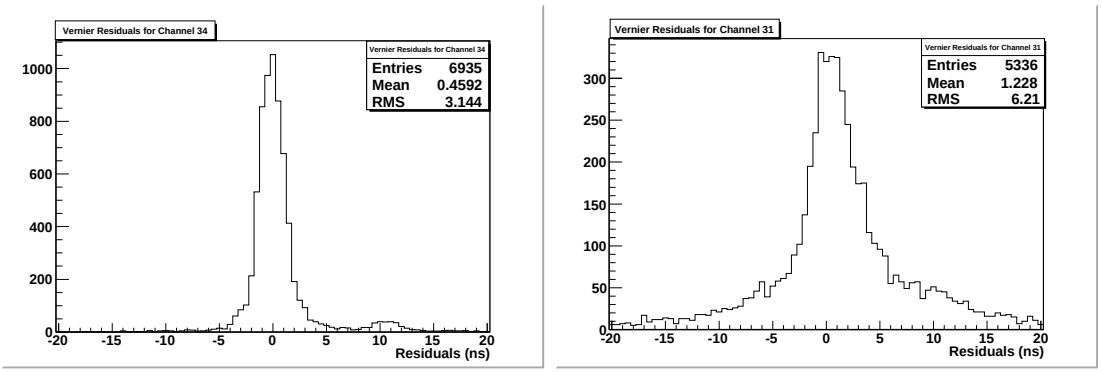

Figure 6.1: Distribution of vernier times for two different channels over the course of a typical astronomical data run, after correction for the vernier peak. The channel on the left corresponds to a heliostat near the centre of the array, that on the right belongs to a heliostat far from the tower and on the edge of the array. The peaks close to zero offset from the expected wavefront (at 0 ns) are from Cherenkov light showers, whereas the uniform background is due to random hits from NSB. The lower occupancy and higher susceptibility to random hits in the case of the outlying heliostat is evident.

All calculated quantities are stored in separate `stoff` banks, which are written to an output file along with some of the original banks. Only part of the FADC traces (a window of 48 ns around the trigger time) are transferred to the `Pass0` output in a

calibrated form. Raw data banks are not used in the further analysis, and are hence dropped after the data reduction with `Pass0`.

6.3 Field brightness corrections

During the early stages of analysing data from STACEE, it became apparent that the actual number of triggers from gamma-ray initiated air showers was very low compared to the fluctuations in the trigger rate due to changes in night-sky background, even for the brightest gamma-ray sources. Naturally, night-sky background (henceforth *NSB*) light alone would not be able to trigger a Cherenkov light telescope such as STACEE, as it is incoherent. However, the level of NSB does affect the probability of a given Cherenkov light shower triggering the detector. In the following it will be examined why this is the case, in which situations it becomes a problem, and how it can be corrected for.

6.3.1 Promotion effects

The signals from STACEE's PMTs were AC-coupled before further processing. While an increase of the NSB photon rate would therefore not lead to higher DC fluxes, it did increase the waveform variances. The superposition of NSB fluctuations and Cherenkov photons from a shower that would normally be too dim to cause a channel's waveform to exceed the set PMT threshold could possibly cause the combined waveform to be "pushed" across threshold, in an effect called *promotion*. For obvious reasons, the trigger promotion effect becomes more pronounced with increasing NSB fluxes, which are reflected in the PMT rates and the PMT currents (as measured before the high-pass filters).

One may argue that for sufficiently thin Cherenkov wavefronts, equal number of showers are promoted as are lost due to negative NSB fluctuations at the exact position of the wavefront peak. However, as we will see from the following, even in this case promotion exceeds "demotion" by far. STACEE's trigger rate was almost exclusively due to hadrons even for bright gamma-ray sources, and hence the vast majority of the Cherenkov light showers available to be promoted were also of hadronic origin. Since the cosmic-ray spectrum falls off like a power law with spectral index $\alpha \simeq 2.6$ - 2.7 (see Sec. 5.5.5), the likelihood of a sub-threshold – *i.e.* low-energy – event being promoted was much greater than that of a higher-energy event being prevented from reaching the trigger threshold. The net effect of increasing NSB rates was therefore an increase in trigger rates.

Promotion in itself is only a problem for ACTs when energy spectra are to be determined, but it does become an issue even in the ON–OFF analysis whenever the NSB levels differ between ON and OFF runs of the same pair. Depending on whether the NSB flux is higher in the ON or the OFF run, the result may be an artificial excess mimicking a gamma-ray signal, or the masking of a real gamma-ray excess. This situation may occur as a result of deteriorating or improving sky conditions, but usually it happens because the ON and OFF sky regions in the fields of view of the heliostats have different overall brightnesses. In the absence of temporary effects such as reflecting clouds, the latter are determined exclusively by the stars in the fields of view of the heliostats. It should be noted that while all ACTs are affected by the promotion effect, a detector with a large field of view and limited hadron-rejection capabilities such as STACEE is particularly susceptible to artificial biases resulting from an unequal brightness from the two fields of view.

In the absence of any way to measure the expected hadronic promotion rate for a given (ON–OFF) field brightness difference, reflected in two sets of PMT anode current values (one each for ON and OFF source observations, respectively), one may use simulations of hadronic showers to predict the trigger rates. Fig. 6.2 shows ON/OFF trigger rate ratios obtained by running a large suite of simulated hadronic showers, *i.e.* air showers due to proton and helium primaries, multiple times though the STACEE simulation chain (see Chapter 5). Each data point corresponds to the conditions of a real source observation pair, in particular in terms of the PMT currents and rates. The linear dependence of the promotion trend on the average current difference is evident. As we will see later, simulations depend very sensitively on a large number of parameters, and thus a fine tuning of all these parameters is essential in order to accurately mimic the conditions pertaining to a given observation. At this point we will therefore only use simulations to elucidate qualitative trends, in particular since a proper equalisation of the background rates is done using a different method, as we will see in Sec. 6.3.2.

In order to quantify the effect of event promotion experimentally, STACEE observed several stars of different magnitudes, which have no known gamma-ray sources in their fields of view and whose background (OFF) fields are sufficiently dark. Fig. 6.3 shows the resulting excess rates as a function of the average PMT current differences between ON and OFF runs. The excess rate is defined as

$$R_{\mathrm{excess}} = \frac{N_{\mathrm{ON}}}{t_{\mathrm{ON}}} - \frac{N_{\mathrm{OFF}}}{t_{\mathrm{OFF}}} \tag{6.1}$$

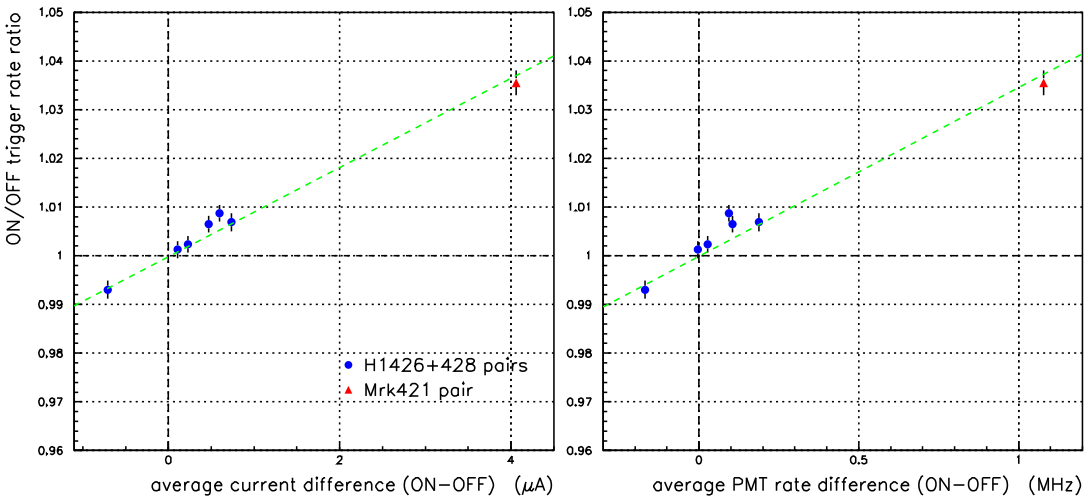

Figure 6.2: Expected ON to OFF trigger (L2) rate ratios due to measured differences in PMT currents (*left*) and rates (*right*) of several representative source pairs (data from 2003), according to simulations. The same suites of simulated hadron showers from the transit points of the sources H 1426+428 (*blue circles*) and Mrk 421 (*red triangle*) were used for both ON and OFF, under identical conditions. The trigger rate ratios are plotted as a function of the average currents and rates. The dotted green lines are linear, zero-constrained fits to the data points. Errors are statistical.

where N_{ON} and N_{OFF} are the total number of events observed in all ON-source and OFF-source runs for each source, with corresponding accumulated live times t_{ON} and t_{OFF}.

In the absence of any other means of correcting for the different background light levels, actual gamma-ray rates can hence be estimated from observed excess rates by calculating the offset from the baseline fit shown in Fig. 6.3, *i.e.* by subtracting the expected rate due to hadronic promotion given by the linear fit. Naturally this method is rather crude, not only due to the very limited accuracy of the fit (note the large error bars of the data points) but also due to the fact that a given value of the average current may be the result of many different channel-wise distributions of the currents. A more accurate way of correcting for the field brightness differences is thus desirable.

6.3.2 Software padding

From the arguments presented in the preceding section, it is immediately clear that correcting for the promotion effect is an essential prerequisite for any meaningful interpretation of STACEE data. Fortunately, the recorded FADC traces represent a powerful tool for accomplishing this. The idea behind the technique called *software padding*, which

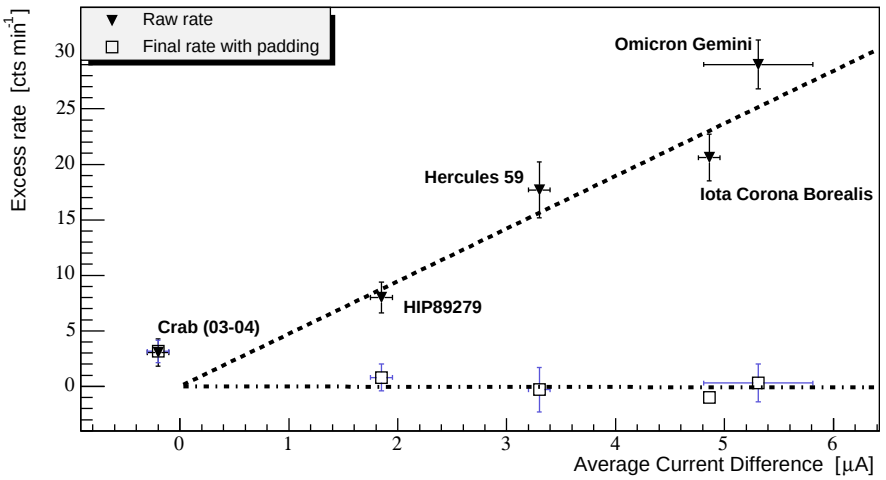

Figure 6.3: Excess L2 trigger rates as a function of average PMT anode current differences between ON and OFF observations, for several stars of different magnitudes. The dashed line is a linear fit constrained to go through the origin, with a slope of (4.74 ± 0.29) min^{-1} μA^{-1}, thus representing STACEE's zero-signal baseline. For comparison, the raw excess rate of STACEE's observations of the Crab Nebula (2003-04) is also shown; note the negative current difference and the excess above baseline which is due to gamma-ray events. The open symbols correspond to the final rates of the various star data sets *after* padding; they are compatible with zero (dot-dashed line). Adapted from Lindner (2005).

is widely used among the gamma-ray astronomy community, is to add artificial noise to the recorded traces of whichever half of the pair (ON- or OFF-run) has a lower level of intrinsic NSB noise. The fundamental principle behind the technique is that of linear superposition, *i.e.* the fact that any two AC-coupled traces can be superimposed to create a summed trace which will have the same characteristics as the two original traces, provided the latter contain the same Fourier components.[2] As the PMT anode currents increase, more single-PE pulses are added to the FADC traces, resulting in an increase of the RMS-value. Thus in order to compensate for the effects of different NSB fluxes, the RMS-values of the corresponding traces need to be equalised. As long as the components with which the traces are built up do not change, we know that:

- the magnitude of the promotion effect is a direct function of the RMS ($\sqrt{<\sigma^2>}$)
- the variance σ^2 is directly proportional to the PMT anode current (see Eq. 4.4)

[2] In other words, as long as they are superpositions of the same basic elements, in this case identically shaped single photoelectron pulses.

When implementing the technique, one needs to remember that RMS-values add in quadrature, and we therefore need to add the *variances*. This is true as long as the two distributions are Gaussian, which is approximately the case here. If $\sigma^2_{\rm hi}$ is the variance of a trace in the noisier half of a pair and $\sigma^2_{\rm lo}$ that in the quieter half, equalisation of the RMS values requires adding a *padding* trace of variance $\sigma^2_{\rm pad}$ to the less noisy trace so that

$$\sigma^2_{\rm hi} = \sigma^2_{\rm lo} + \sigma^2_{\rm pad} \tag{6.2}$$

In practise, this is done by means of the trace variances stored with the FADC data, calculated at data acquisition time from the 400 ns of trace before the arrival of any signal photons (see Sec. 4.4.5). Naturally, the ON/OFF variances of each of the 64 STACEE channels are equalised separately. Furthermore, since background light levels have been observed to change drastically even within one and the same run, each run is divided into intervals of 26 seconds within which the variances of individual traces are averaged. The traces that are to be added, *i.e.* of variance $\sigma^2_{\rm pad} = \sigma^2_{\rm hi} - \sigma^2_{\rm lo}$, are part of a library of 1000 traces which were recorded in special runs, using an LED as a variable source of background light. This *library padding* scheme is a simple, yet robust way of correcting for the variance differences between ON and OFF data. Of course, care must be taken that the runs used for creating the library traces were not contaminated by excessive amounts of RF noise or other electronic artifacts, and that any channel-to-channel differences in regard to the single-PE pulse shapes were taken into account.

While equalisation of the ON and OFF trace variances does equalise the software re-trigger probabilities of showers of the same energy within both halves of a pair, it does not recover showers in the less noisy run that failed to trigger in the first place, but would have passed the trigger requirements if they had been promoted by the extra NSB. The only way to "level the playing field" completely is by subsequently re-imposing the trigger condition at a higher threshold in both halves of the pair. The complete padding cut therefore consists of padding the traces of the dimmer half-pair, as well as re-imposing the hardware trigger condition with all thresholds raised by an amount X. The trigger re-imposition is done by means of a *static gate cut*, in which only hits located within a time window of 16 ns width centered on the dead-reckoned shower front position are considered by the software trigger logic.

Finding the amount X by which the thresholds need to be raised in order to purge all excess triggers in the brighter half of the pair is not intuitive. One straightforward way of dealing with this problem is by determining the value of X at which the promotion-related excess in star data vanishes. This approach however is problematic because the

obtained value of X will depend on the average current differences of the particular star observations. Likewise, one fixed value of X can only be appropriate for one particular value of the average current differences between ON and OFF observations; it may be too high for observations with very small brightness differences of the background fields, causing the thresholds to be raised by unnecessarily large amounts and thus the loss of many low-energy gamma-ray events. For pairs with very high NSB differences, on the other hand, it may be too low, preventing the padding procedure from completely removing the promotion bias.

An *ad hoc* way of finding the right threshold increase X is by choosing a star with similar current differences as the source observations to which the padding cut is to be applied. As STACEE did not observe nearly enough stars to actually parameterize this relation – a statistically significant sample of star data (5 - 10 ksec live time) required many ON/OFF pairs – the user analysing STACEE data is left with the choice of either to be very conservative and choose a high value of X, or to be more "aggressive" and estimate an appropriate value of X from the available star data sets. As a representative example, Fig. 6.4 shows the final rate (*i.e.* the excess rate after all time cuts, see Sec. 6.4) of STACEE observations of the star *Hercules 59* as a function of the amount X by which the PMT thresholds were raised in the offline analysis. We can see that the excess, which is entirely due to promotion of hadronic showers, vanishes for values of $X \gtrsim 15$ mV.

Both the variances for each run, channel and padding time interval, as well as effective thresholds for each channel and run were calculated beforehand and are stored in a

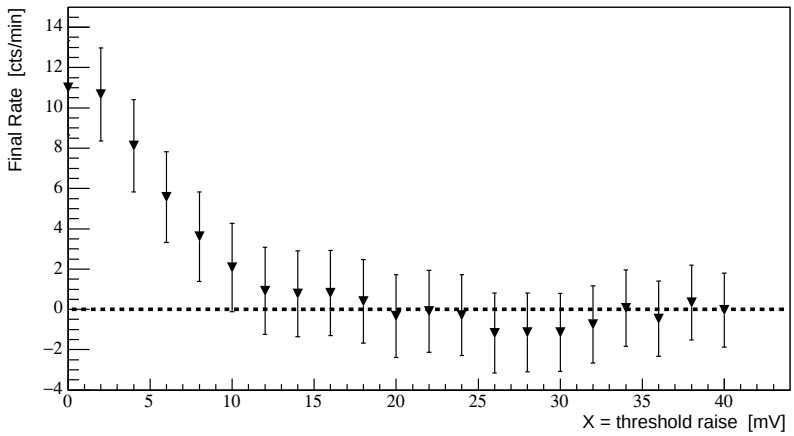

Figure 6.4: Final excess rate after all time cuts for observations of the star *Hercules 59* ($m_{\mathrm{B}} = 5.29$) in the summer of 2004, as a function of threshold increase X. Heliostats were paracanted. The average current difference for this star was $\Delta I = (3.7 \pm 0.2)\,\mu$A, the raw excess rate was (22 ± 3) min^{-1} (Lindner, 2006b). Adapted from Lindner (2005).

mySQL database. This allows the software padding to be done within `Pass0` on a run by run basis, *i.e.* without simultaneously examining the run corresponding to the other half of the pair. The re-trigger thresholds are determined by adding the required threshold increases X to the effective thresholds of each channel.

6.4 Data-quality cuts

In the previous section we have described how the raw STACEE data are refined and converted into a basic calibrated form by processing each run individually. Because of STACEE's very limited angular resolution, any meaningful interpretation of the data requires every on-source run to be paired with a corresponding off-source observation. The next step in the analysis of STACEE data is hence to examine both halves of a pair together, with the aim of extracting all usable sections of the data, *e.g.* by removing temporal segments that are problematic in either ON- or OFF-run. Since these kinds of cuts are based on time and affect both halves of the pair, they are also called *time cuts*. By *usable* we mean, in this context, sections of on-source runs for which usable OFF observations exist (*i.e.* acquired at the same local sky coordinates), and which are not affected by any of a number of possible types of data corruption, which we will examine later. There also needs to be a convenient way in which *event cuts*, which are based on parameters calculated by either the basic (`Pass0`) or the advanced (`Pass1`) analysis programs, can be applied to individual events. Event cuts include the padding cut (Sec. 6.3.2) and any hadron-rejection cuts (Sec. 6.6).

This stage of the analysis is taken care of by a program called `stoff_pair`, which reads in both on- and off-halves of a pair simultaneously. Being a modular program, it contains a framework that performs basically all the necessary steps towards a complete analysis of the data, including time cuts, event cuts and the calculation of ON-OFF significances and final excess rates. Among the basic tasks that are performed are ensuring that both run files are indeed part of the same observation (*i.e.* form a valid pair), and aligning the start- and stop-times of both halves of the pair. Only overlapping parts of a pair are used in further analysis. The types of time cuts that are applied to the data within `stoff_pair` are described below.

6.4.1 Heliostat malfunction cuts

Heliostat tracking errors or mispointings of heliostats can seriously degrade the data quality. This is particularly detrimental if a heliostat performed well in one half of the

pair, but was experiencing a problem in the other half. Malfunctioning heliostats may lose the ability to collect Cherenkov photons, and generally result in erratic changes of background light levels, affecting the probability of promotions of hadronic events and possibly leading to accidental triggers.

During data analysis, the performance of individual heliostats can be evaluated by examining the heliostat status information which is included in the data stream (see Sec. 4.1.1). The latter includes the actual headings of all heliostats recorded at periodic time intervals of 20 seconds, as well as a number of error flags. If one of the error conditions was met, *or* if a heliostat's pointing was more than 2 bits (about 0.1°) off in either azimuth or elevation, the corresponding channel was flagged as *bad* for the duration of the two adjacent 20-second time intervals. Rather than removing any sections of data which had one or more heliostats flagged as erroneous, only time intervals for which there were *different* heliostats malfunctioning during the ON and OFF observations are cut. This less strict criterion was introduced to prevent the loss of excessive amounts of data due to heliostats which experienced consistent problems during the whole observing night and had therefore been stowed or deactivated prior to observing, or isolated heliostats with constant minor tracking problems.

6.4.2 Frost and dew cuts

A rather subtle, yet fatal problem for STACEE was water condensation on the heliostat facets in the form of frost or dew. Condensation takes place whenever the air temperature cools down to the *dew point*, defined as the temperature at which the humidity content of the air is at saturation (100%). On some objects with high radiativity and low heat capacity such as STACEE's heliostats, frost or dew can form even at temperatures slightly above the dew point. Despite the generally dry climate at the observation site, frost occurs occasionally during the winter months (November - April), and dew is possible also during wet spells in the spring and fall.

The effect of frost and, to a lesser extent, dew on the heliostat mirrors is a diffuse, semi-isotropic scattering of Cherenkov light and ambient background light, resulting in a severely degraded transmission for Cherenkov light, while background light levels are not affected or may even increase. A gradual buildup of condensation thus usually made itself known in the form of steadily declining L2 trigger rates. Runs affected by frost are unusable, mainly because frost coverage seldom stays constant but rather increases with time as the night progresses and the temperature drops. As a result, ON and OFF runs

taken during frosty nights have radically different optical throughputs and L2-trigger rates.

The only way to unequivocally detect condensation was to inspect the heliostats visually by walking into the heliostat field. What can be done offline, however, is to cut out all observing time for which the weather station data indicate conditions favourable for condensation. The quantity to which the cut is applied is the difference between temperature and dew point, *i.e.* the frost index (see Sec. 4.3.2). Whenever this value is below 10 degrees Fahrenheit, condensation is considered possible. For this analysis, all data with a frost index of less than 10 F were hence discarded.

6.4.3 L1-rate cuts

Bad weather seriously affected STACEE's duty cycle. While the detector was rarely deployed during cloudy nights, whenever sky conditions were marginal and it had cleared up enough for Cherenkov observations to appear possible, many data pairs later turned out to be unusable because of drastic current fluctuations caused by high, barely visible clouds or low, quickly moving clouds that were illuminated by stray light from the city of Albuquerque. A means to detect bad weather, or more generally unstable observing conditions, is therefore essential. Although STACEE's atmospheric monitor system (Sec. 4.3.2) turned out to be of little use, fortunately the Cherenkov light data itself contained a quantity that was very sensitive to changes in sky conditions, namely the L1 (cluster) rates. Due to the fact that the cluster rates were driven by random coincidences and depended very sensitively on the rates of the eight contributing PMTs, very slight changes in the latter due to increased background light levels translated into large excursions of the cluster rates (recall Eq. 4.6). By monitoring the L1 rates, the DAQ operator was able to detect deteriorating sky conditions, visible in the form of rising and/or rapidly fluctuating rates, in real time. Examples of L1 rate plots for two representative source pairs are shown in Fig. 6.5. Note the vastly different magnitudes of individual clusters' rates, which is caused by the exponential dependence of the L1-rates on individual PMT rates.

In the same way as they were useful for monitoring the incoming data, the L1 rates can be employed to apply time cuts. Pairs for which either ON or OFF run were marred by severely fluctuating L1 rates, indicating the presence of clouds, are discarded and not analysed any further. All other pairs, including those which exhibit reasonably stable L1 rates for at least part of the time, are examined using an algorithm that cuts out sections of data according to criteria which will be discussed below.

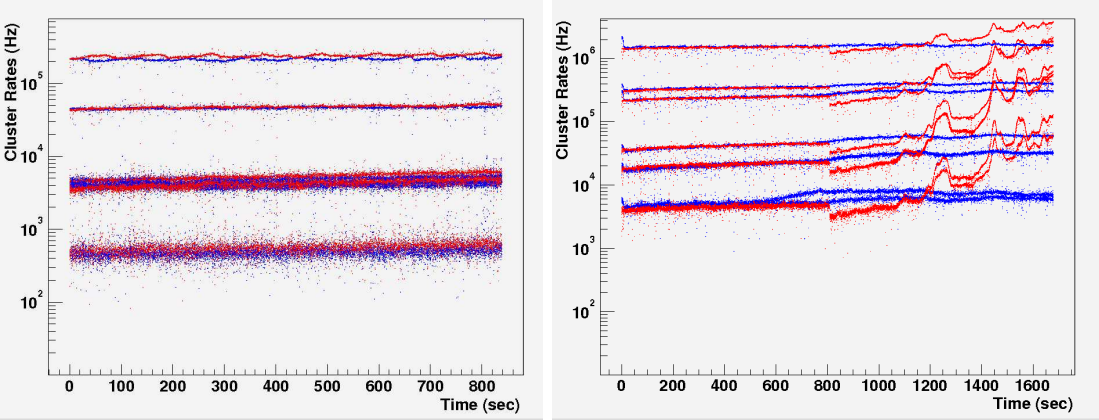

Figure 6.5: Superimposed on-source (*blue*) and off-source (*red*) L1 rates for two different source pairs (*left*: runs 11123/11124, *right*: runs 12114/12115). The pair on the right is marred by a global rate drop, presumably to an ambient RF source being turned off, and a subsequent *rate hike* due to increasing cloud cover.

Experimenting with different algorithms showed that a reliable automated procedure for cutting sections of data with so-called "unstable" L1-rates is not straightforward; frequently, steadily rising or falling L1-rates (*e.g.* due to passing clouds), or intervals with small variations but abnormally high rates were mistaken for clear-sky conditions. Hence, a procedure that used the L1-rate *ratios* between OFF and ON data as cut parameter was eventually implemented, thereby introducing a means to cut on the relative values of the rates. Using the ratio of the rates as a descriptor instead of absolute values for ON and OFF cluster rates is necessary because the latter are not constant, even during "good" conditions. The main variables upon which the L1 rates depend are:

- the heliostat target position, in particular elevation, because the amount of background light scattered by the atmosphere increased with increasing zenith angle
- the exact values of the PMT gains, which were expected to drift somewhat within each observation night as the PMTs warmed up, and also gradually over the course of the observation campaign
- the degree of sky clarity, determined by moisture and aerosol contents
- malfunctioning or inactive channels, or channels with high RF-noise amplitude

Independent of the actual L1 rates, data taken in stable conditions are expected to show a good correlation between ON and OFF L1-rates for all observations of a given source, while unstable conditions in general give rise to uncorrelated ON and OFF rates.

The actual cut procedure consists of the following steps. The whole set of pairs taken on our prospective gamma-ray source of interest are divided up into pairs for which the OFF observation preceded the ON observation (OFF-West/ON) and ones where it followed (ON/OFF-East). The L1-rate cut program is then run on each set separately. This is important, because in general the background star fields of the two off-source positions are different in overall brightness as well as in morphology. For the same reason, all pairs of non-standard lengths (*e.g.* short pairs of 14 minutes) need to be processed independently. The L1-ratio cut program first reads through all the pairs once, but ignoring sections of "bad" data that have already been cut by either of the two previously described time cut programs. The L1 rates of the "good" sections are averaged over intervals of 30 seconds, and scatter plots are filled of off-source *vs.* on-source rates for each cluster, such as the ones shown in Fig. 6.6. The linearity between the (logarithms of the) ON and OFF cluster rates for the bulk of the data in these plots demonstrates that the assumption about a constant ratio between ON and OFF rates is correct. Sections of the data where the linearity condition is not fulfilled should consequently be removed.

In order to identify and cut these intervals, the data are histogrammed in the form of $\ln(R_{off}/R_{on})$, where R_{off} and R_{on} are the L1 rates of the OFF and ON run respectively. This is approximately the same as making an angular projection of the scatter plots, and results in almost symmetric, peaked distributions like the ones shown in Fig. 6.7. While the peaks are not really Gaussian in shape – note for instance the slight asymmetry which is a result of the projection method – they can be approximately described by a Gaussian peak, thought to be due to intervals which have the expected L1-rate ratio, and long tails on either side of the peak caused by data quality problems in either half of the pair. A Gaussian is thus fitted to the peak area, not including the tails, and the actual L1-ratio cut is centered on the fitted Gaussian. Different cut widths have been used within the STACEE collaboration; for this analysis, a cut of 3 standard deviations ($3\,\sigma$) was chosen, since the default cut at $2\,\sigma$ would have removed some good quality data. Gaussian fits and $3\,\sigma$-cut windows are shown on the plots in Fig. 6.7.

While the pair on the left of Fig. 6.5 passed the L1-ratio algorithm in its entirety, the second half of the pair on the right was cut due to the fluctuating and rising L1 rates in the off-source run, indicating increasing cloud cover.

Atmospheric instabilities are not the only problems detected by the L1-scatter algorithm; non-weather related disturbances such as transients due to bright lights from airplanes passing through the field of view of some heliostats, or headlights from vehi-

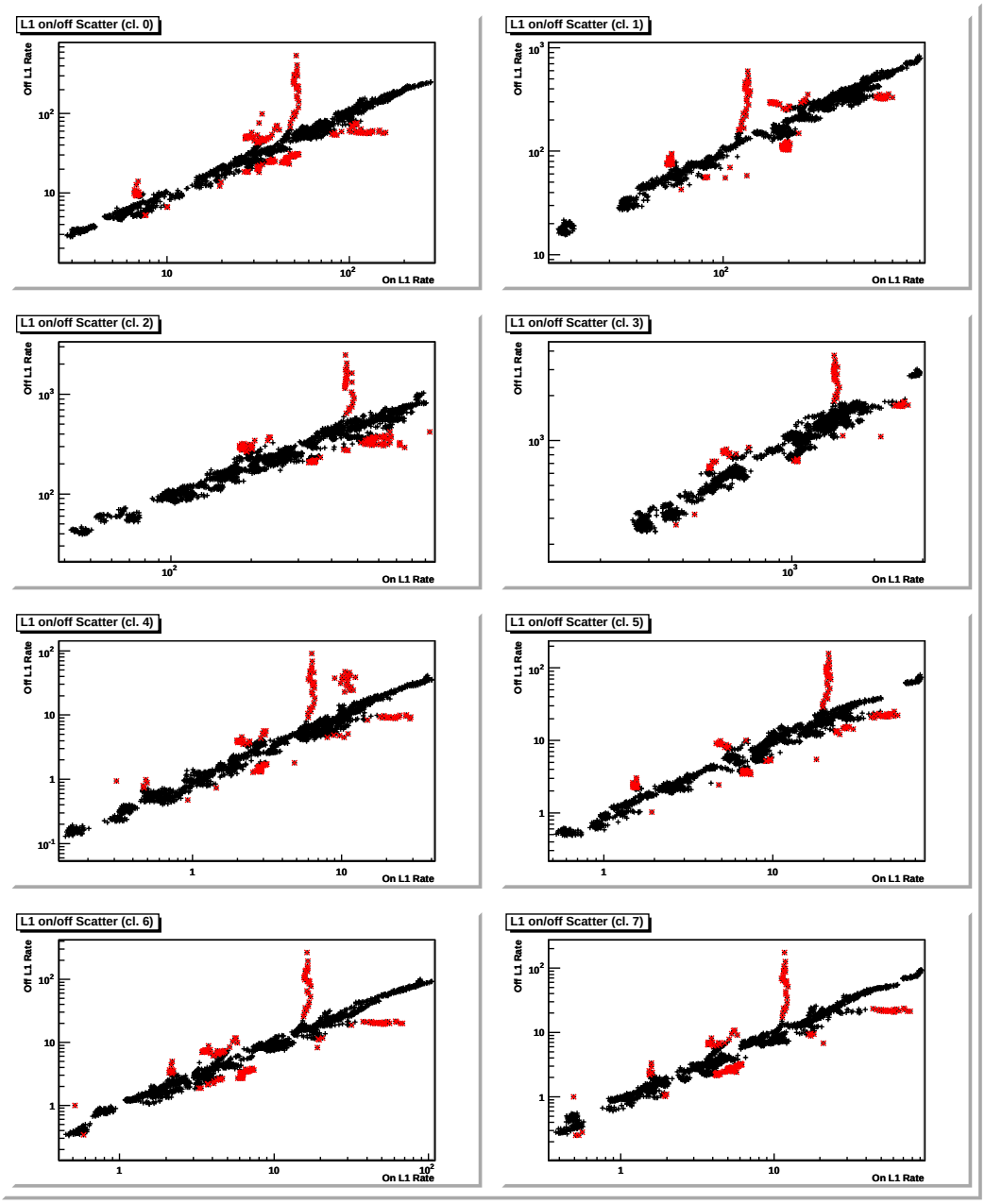

Figure 6.6: Scatter plots of OFF *vs.* ON cluster rates for all 8 clusters, containing all usable sections of pairs with the OFF field to the East of the ON field, for the 2004 observations of H 1426+428. Each point represents a time interval of 30 seconds. The points shown in black passed the L1-ratio cut of width 3σ, the ones in red did not. The slightly differing ranges and slopes of the distributions of individual clusters originate in the different fields of view of the heliostats of each cluster in the sky, caused by the canting of the heliostats.

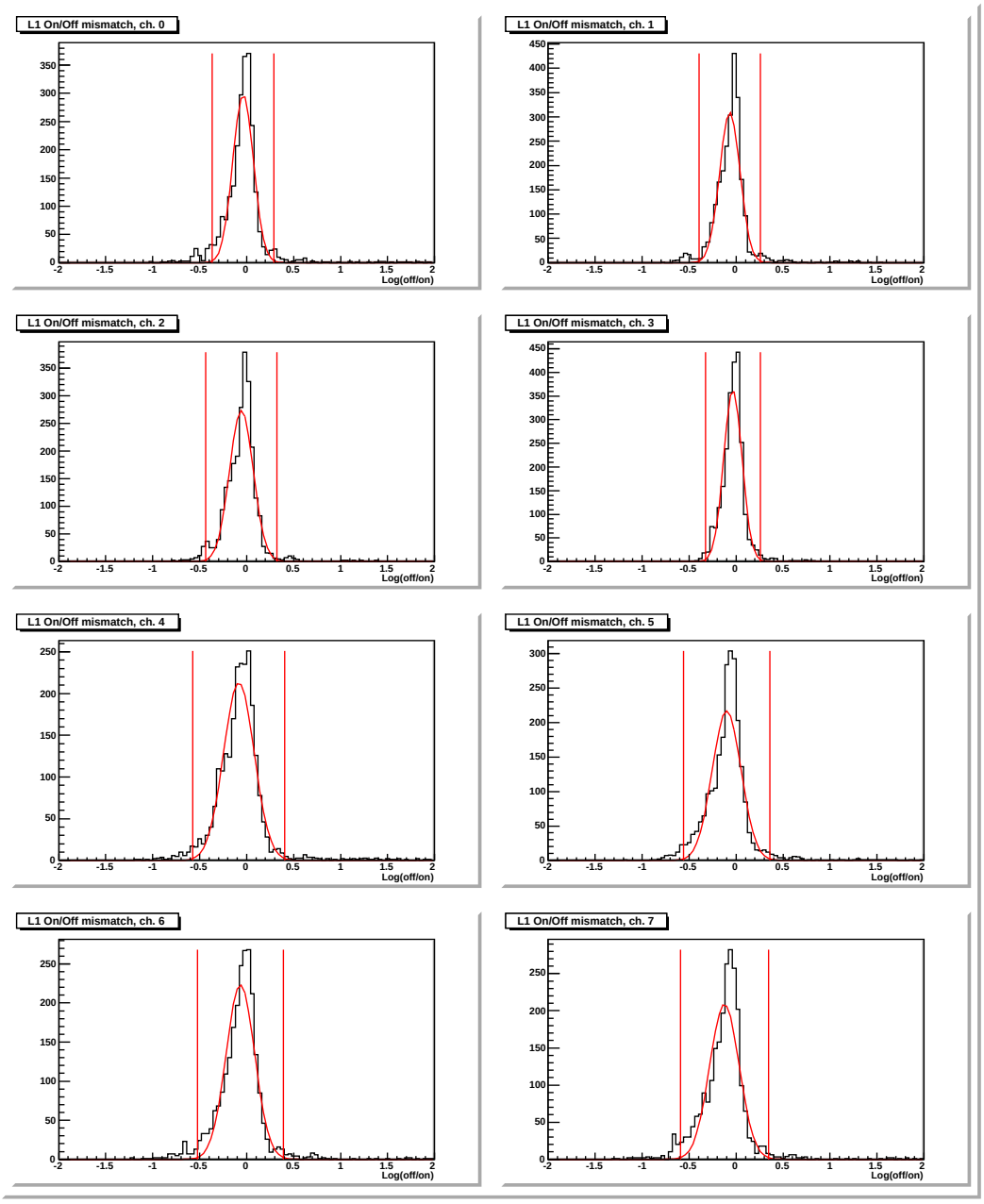

Figure 6.7: Distributions of the logarithms of OFF/ON L1-rate ratios for all 8 clusters, containing all usable sections of pairs with the OFF field to the East of the ON field, for the 2004 observations of H 1426+428. Gaussian fits to the peaks and the cut window at $3\,\sigma$ from the fitted peak positions are shown in red.

cles, are also removed whenever they gave rise to spikes in the PMT rates of at least one cluster.

6.4.4 Other detector-related cuts

Apart from the fundamental data quality cuts described above, there are a few other time cuts which are subsequently applied to remove sections of data where parts of the detector itself did not operate properly:

- Since all of the event reconstruction and advanced data analysis relies on the FADC data, sections of a pair which are missing the data from at least one of the four FADC crates were removed from further analysis. This condition was usually caused by the FADC-DAQ program crashing on one or more of the corresponding computers. Even though the custom-written real time Linux driver could support STACEE's highest event rates under normal conditions, it occasionally failed and caused errors, among these an abnormal termination of the run.

- Any sections of data for which the padding process failed was cut out. Normally, data which has FADC information for all 64 channels can be padded successfully, however in isolated cases the recorded variance differences exceed the dynamic range over which padding is possible. This may be due to excessive (RF or NSB) noise, or errors in the variance calculation. Due to the importance of padding for an unbiased analysis, these sections need to be removed.

- Any sections of time during which one or both of the high-voltage supplies were turned off was removed from the data. This situation occurred *e.g.* when the HV was automatically shut down by the DAQ due to a current surge, in order to protect the PMTs. Sudden PMT current increases were usually due to airplanes passing through the fields of view of some heliostats, or headlights from vehicles.

6.5 Detection Significance

The output of the various `stoff_pair` time cut programs described in the previous section consists of user-friendly ASCII files, which can be concatenated to yield a final cut file containing all the time cuts desired for the further analysis. The purpose of the time cuts is only to select a clean sample of all available data by removing faulty or biased sections from the data set. For more complex cuts, those data sections which passed the time cuts need to be examined event by event. These so-called *event cuts*

are used to select or reject certain events, with the ultimate goal of keeping as many gamma-ray events and rejecting as many hadronic events as possible, thereby increasing the chance of detecting a gamma-ray signal. These types of cuts will be introduced in Sec. 6.6.

Even without event cuts, the cleaned and software-padded data set can already be examined for an excess of events in the ON run. Due to the limited statistics by which gamma-ray astronomy is plagued – even for the new generation of imaging ACT – the standard practise is to associate a *significance* (S) with each observation, independent of whether the observation leads to a detection, merely an on-source excess or even a deficit. As suggested by the name, the significance is measured in units of standard deviations (σ) by which the measured signal exceeds the background.

6.5.1 Calculating the significance

For a given observation, characterised by an observed number of events N_{ON} in the on-source run and N_{OFF} in the associated off-source run, with corresponding live times t_{ON} and t_{OFF}, the significance can be calculated in a straightforward manner by using the ratio of the two live times to scale the event numbers:

$$S = \frac{N_{\mathrm{ON}} - (t_{\mathrm{ON}}/t_{\mathrm{OFF}})N_{\mathrm{OFF}}}{\sqrt{N_{\mathrm{ON}} + (t_{\mathrm{ON}}/t_{\mathrm{OFF}})^2\, N_{\mathrm{OFF}}}} \tag{6.3}$$

While the above equation provides an intuitive way of calculating the detection significance, large live-time differences between on- and off-source observations require a different way of scaling the count rates. A slightly more accurate way of obtaining S under application of maximum-likelihood methods is given by Li & Ma (1983), which is currently the standard way of calculating detection significances within the gamma-ray astronomy community. The error made by using Eq. 6.3 is however quite small; for the very small live-time differences typical for STACEE, it is less than 1%, and hence negligible.

For ACT, which due to the nature of the atmospheric Cherenkov technique have rather low signal-to-background ratios, a single observation pair by itself has usually no statistical power. For STACEE, all observations performed in a given observing period are therefore summed, and a cumulative significance is calculated from the summed on- and off-source event numbers and the corresponding live times. It is often convenient to remember that the significance parameter as defined in Eq. 6.3 is related to the statistical error $\Delta R_{\mathrm{excess}}$ of the measured excess rate (see Eq. 6.1) via $\Delta R_{\mathrm{excess}} = R_{\mathrm{excess}}/S$.

6.6 Event Reconstruction

All data analysis methods described up to now are only concerned with isolating clean and equalised sections of overlapping on- and off-source observation time, with the aim of searching for an on-source excess above background. As mentioned previously, STACEE's sensitivity is quite low when restricted to using only the ON/OFF technique to detect gamma rays.[3] In order for the experiment to become more competitive and produce meaningful results, the raw sensitivity clearly needs to be improved. From data taken with a given, fixed hardware configuration, this goal can be achieved by rejecting a large fraction of the hadronic background events during offline analysis, while at the same time retaining as many gamma-ray events as possible. In other words, the quantity which is to be maximised is the quality factor Q (Eq. 4.1), which we shall write here in the form

$$Q = \frac{N'_\gamma / N_\gamma}{\sqrt{N'_{\rm hadron}/N_{\rm hadron}}} \tag{6.4}$$

where N_γ and $N_{\rm hadron}$ are the total numbers of gamma-ray and hadronic events in the data set, respectively, and the primed quantities denote the respective numbers of events that pass a given event cut.

Unfortunately, the standard methods by which imaging ACT are able to reject a large portion of hadronic events (see Sec. 3.2.1) cannot be applied to wavefront-sampling detectors such as STACEE, as they rely on imaging Cherenkov showers in angular space. The latter type of experiments must therefore use different techniques to reject hadronic background. Despite their very limited *angular* resolution, wavefront samplers have one advantage over conventional, single- or few-dish imaging Cherenkov telescopes: the large cross-sectional area on the ground across which they are able to sample the Cherenkov wavefront, which, when combined with a good (~ 1 ns) timing resolution, allows for a *spatial* and *temporal* characterisation of the wavefront. As we have seen in Sec. 3.1.3, the spatial-temporal Cherenkov light distributions of gamma-ray induced air showers differ significantly from those typical of hadronic showers. The fitting of wavefront parameterizations to measured arrival time and charge distributions of all 64 STACEE channels can hence be used to obtain likely values of the shower core locations and shower energies; at the same time, the goodness of such fits can be used as a discriminator between gamma-ray and hadronic showers.

The other application that requires a reconstruction of individual events is the measurement of energy spectra. While the mere detection of a gamma-ray source allows

[3] Using only raw excess rates, two full observing seasons were required to accumulate a $5\,\sigma$ signal of gamma rays from the Crab Nebula.

for the calculation of an integral flux, spectral measurements are scientifically of much greater value, since they provide a far more accurate description of a measured gamma-ray emission spectrum. In order to measure an energy spectrum, individual gamma-ray candidate events need to be assigned an energy value. Naturally, the ability of obtaining spectral information from observations of a given gamma-ray source depends on the sensitivity and energy reconstruction capabilities of the ACT. In STACEE's case, the latter two are only sufficient for obtaining spectral information for the brightest gamma-ray sources, such as the Crab Nebula and Mrk 421 (see *e.g.* Carson *et al.*, 2007). For sources with lower fluxes, the number of registered gamma-ray events is generally too low to allow for sufficient statistics in individual energy bins. Despite the fact that no energy reconstruction is performed as part of this work, the principle of event (and therefore energy) reconstruction shall be summarised in the following sections, as it is relevant for understanding the basics of gamma-hadron separation for STACEE. While the former methods have been developed over the years, the emphasis will be on a second, more recent technique, which is also used for the data analysis presented by this thesis.

6.6.1 Charge template method

All advanced reconstruction methods for STACEE take advantage of the detector's large physical cross-section, which allows it to sample a substantial fraction of the Cherenkov light pools of all triggering air showers. In the so-called *charge template* method, detailed descriptions of which can be found in Scalzo (2002b,c) and Scalzo *et al.* (2003), simulations of gamma-ray showers are used to build up large look-up tables (the "templates"), consisting of a number of expected photoelectrons on each heliostat (*i.e.* channel) as a function of the detector heading, primary energy and shower core position. During data analysis, the measured dead-reckoned charges in all 64 channels for a given event are compared to the templates, and a maximum likelihood fit is performed in order to determine the shower core, and subsequently the shower direction and energy.

Core reconstruction

The first step towards a successful shower reconstruction is to locate the shower core. In the initial direction fits mentioned in Sec. 6.2, the shower core was assumed to be the centre of the heliostat array, *i.e.* at the location for which the trigger probability is highest. As one may understand when looking at Fig. 6.8, there exists a degeneracy between shower core position and primary arrival direction, which cannot be resolved only with a spherical direction fit. For such fits, either the shower core or the shower

direction need to be fixed in order to obtain the other respective quantity. The shower core is needed as an input parameter not just for direction fits, but knowing it is also essential for any reconstruction of the shower energy. This is because showers of the same energy deposit different amounts of Cherenkov light in the PMTs depending on where exactly they land; in general, more charge is collected from showers with cores near the centre of the array than from showers with larger impact parameters.

There are in fact various techniques for reconstructing the shower core. One way of finding the core location, as mentioned above, involves using the charge templates. The spatial resolution attainable with this method however is not optimal for conventional methods of pointing the heliostats (*i.e.* monocanting), as a certain degree of degeneracy remains. A better sensitivity of the template fitting method to the core location can be achieved by implementing a different canting scheme, in which a subset of heliostats are aimed directly at the celestial coordinates of the target (*i.e.* pointed in parallel mode), instead of at the expected interaction position within the atmosphere. The paracanting scheme used by STACEE from the 2003/04 observing season on was introduced to improve the core reconstruction through templates, after simulation studies had shown that the radial distance from the array centre up to which cores could be successfully reconstructed would increase from 60 m to 125 m (Scalzo, 2004). See Fig. 6.8 for a sketch of the two different canting methods used by STACEE.

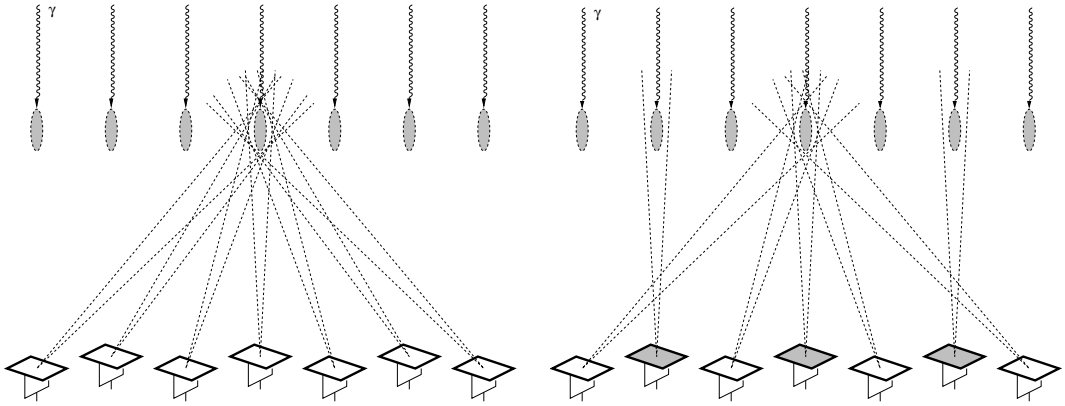

Figure 6.8: Sketches of the monocanting (*left*) and paracanting (*right*) schemes for pointing heliostats. In the former, all 64 heliostats are aimed towards a single point in the atmosphere, the expected position of shower maximum. Paracanting, for which 16 of the 64 heliostats are aimed at the prospective gamma-ray source in direct parallel mode, allows for an improved angular resolution at the expense of a minor loss in sensitivity. The gray ellipses correspond to different locations of shower maximum for gamma rays coming from a source at zenith. (Figures are not to scale.) Adapted from Scalzo (2004).

A second, independent core-finding algorithm involves the Cherenkov pulse arrival times on the heliostats. For air showers from gamma-ray primaries, Cherenkov light emitted by the trailing end of the electromagnetic cascade arrives earlier than light emitted higher up in the atmosphere (recall Sec. 3.1.2). Projected onto the ground, these early photons populate the area intersected by the extension of the shower axis (see Fig. 3.5). A measure of the shower core can thus be obtained by calculating the centre of gravity of the subset of heliostats that yielded the earliest arrival times. Finally, the shower core can be found by means of the gamma-hadron separation technique described in Sec. 6.6.2.

Direction reconstruction

Once a reasonably accurate estimate of the shower core is available, wavefront (*i.e.* direction) fits similar to those mentioned in Sec. 6.2 become more accurate, and can be used for discriminating gamma-ray events against EAS from isotropic hadrons. Wavefronts can be fitted by both spheres and cones; the former provide a rather accurate description of the Cherenkov light distributions of gamma-ray showers at the low energy end of STACEE's acceptance range ($\sim 100\,\mathrm{GeV}$), and the latter fits improve with increasing gamma-ray energy. Simulations show that for gamma rays from a point source, the direction can be reconstructed to an accuracy of about 0.2° for both mono- and paracanting for all but the lowest energies.[4] As the shower energy decreases, so does the angular resolution, for monocanting more quickly than for paracanting. The angular resolution depends also quite sensitively on the elevation of the source, and degrades with decreasing elevation due to the decreasing effective cross-section of the heliostat field.

Energy reconstruction

Due to the direct correspondence of Cherenkov light density on the ground with primary gamma-ray energy, which was discussed in Sec. 3.1.2 (see in particular Fig. 3.8), the gamma ray's energy can be obtained in a straightforward manner from the total amount of charge deposited in the PMTs. For the same reason, the charge templates were filled by simulating gamma-ray showers of only one energy (in the current scheme, for 200 GeV), and they are then scaled by the fitting algorithm. The proper scaling factor and exponent[5] are determined by simulations.

[4] This is thus STACEE's minimum *point spread function*.

[5] The exponent is close to unity in any case.

The derived energy estimates are affected by a number of error sources, the main ones being the uncertainties in optical throughput and PMT gains of individual channels. This is particularly a problem since the templates were generated in units of photoelectrons and hence do not take into account the response of the read-out electronics. Clipping of FADC traces for energetic showers may also be a problem, even though the peaks of saturating pulses can be fitted with an empirical function. The energy resolution achievable with the template method is around 30%.

6.6.2 Grid alignment technique

The reconstruction of individual events via templates, as described in the previous section, has the disadvantage that it relies heavily on simulations. While simulations appear to predict most observable quantities quite accurately, particularly for gamma rays, they do depend on a very large number of input parameters. As we have seen in Sec. 5.5, knowing all of these quantities to good accuracy at any particular time is next to impossible. Many parameters, such as the PMT gains and atmospheric transparency, fluctuate over time. While a detailed event reconstruction clearly requires simulations, there are ways to reject hadronic events based on *observed* properties of the Cherenkov showers alone.

The technique referred to as *grid alignment* is a method for characterising the "hadronness" of a Cherenkov shower which is tailored to wavefront-sampling ACT. Since this technique turned out to be rather effective and robust, it was adopted as STACEE's main – and commonly only – tool for offline hadron rejection. The original algorithm was developed by the CELESTE collaboration (Manseri, 2004), and was subsequently adapted for STACEE (for a more detailed description, see Lindner *et al.*, 2007). Like the wavefront fit, the grid alignment algorithm exploits the sphericity and smoothness of the wavefronts of gamma-ray initiated air showers. In contrast to the direction fits however, the grid alignment method does not require a shower core position as input, but instead it outputs a likely core position. Inversely, the shower direction is an input parameter; all showers are assumed to originate from the prospective gamma-ray source, and a fixed atmospheric depth is used for the calculation of the shower maximum position. Indirectly, the method is thus also an energy cut.

Grid search

The algorithm operates in two steps. First, the core position (or rather, the most likely location of the shower core) is located.[6] This is done by stepping through a square grid of possible core locations in the horizontal plane, centered on the "nominal" location of the shower core used by the delay calculation, *i.e.* the centre of the heliostat array. In the common version of the algorithm, the grid consists of 30×30 points with a square spacing of 15 m. For each point, the TOF delays of all channels are recalculated for the case that the shower maximum were located along the line from the point – *i.e.* the assumed shower core – to the source position, and at the proper height for the given elevation, which is determined by the traversed atmospheric depth. This procedure is equivalent to moving the assumed shower maximum around on an identical, parallel grid, at an altitude corresponding to the nominal shower maximum and offset in direction of the prospective source. Note that as in the original delay calculation, a spherical wavefront with origin at the shower maximum is assumed. Using the new delays, the individual FADC traces are shifted by the time corresponding to the difference between the latter and the original, dead-reckoned delays which were used for triggering and centering the FADC traces. All 64 FADC traces are then summed, and the heights above baseline (H) of the summed pulses are compared with their widths (W) as measured at half-maximum. The grid point with the largest height-to-width ratio H/W, called $(H/W)_{max}$, is assumed to be the location of the shower core.

The physical justification for the approach outlined above is that for a spherical wavefront, choosing relative delays that correspond to the actual shower maximum location – and thus to the actual core – will result in well timed in Cherenkov signals across all channels, giving rise to a sharp peak in the summed trace. If the applied delays are for a sphere centered at a point offset from shower maximum, the sum pulse will be more spread out. Fig. 6.9 shows two such FADC trace sums, one for the case of correctly aligned delays, *i.e.* at $(H/W)_{max}$, and one for a point at a rather large distance (200 m) from it.

Sphericity estimator

Once the shower core is known – or, more precisely, the point at which the H/W distribution peaks – one can proceed to construct a figure of merit for assessing how spherical the recorded wavefront is. In the standard grid alignment procedure, this is done by

[6] A variant of the grid alignment technique does not search for the shower core, but instead uses core positions determined externally by one of the two methods described in the previous section. In that case, no actual grid *alignment* is performed.

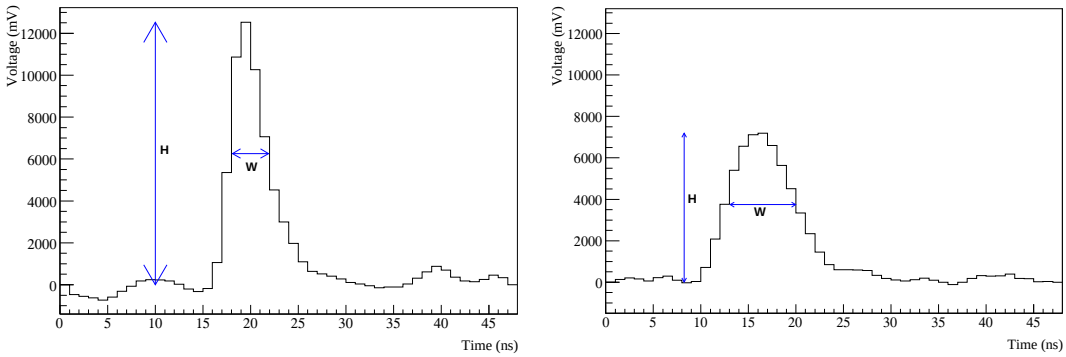

Figure 6.9: Summed FADC traces for a single simulated gamma-ray event. *Left*: Sum at the correct shower core location, for which the TOF delays are properly aligned, maximising the value of H/W. *Right*: Sum for delays corresponding to a point laterally offset by 200 m from the the grid point of $(H/W)_{max}$. The misaligned traces result in a much lower H/W ratio. The ratio of the two ratios is a measure of the wavefront's sphericity. From Lindner *et al.* (2007).

comparing $(H/W)_{max}$ to the average of the H/W values calculated at four background points.[7] The background points are chosen some distance d from the peak of the H/W distribution, offset laterally along orthogonal axes perpendicular to the shower direction.[8] If we denote the average of the background points' H/W with $(H/W)_d$, a figure of merit that describes the sphericity of the wavefront can be defined as

$$\xi = \frac{(H/W)_d}{(H/W)_{max}} \tag{6.5}$$

where smaller values of ξ correspond to more spherical arrival time distributions. By using a cut at an appropriate value of ξ, from now on referred to as *grid-ratio cut*, one should thus be able to reject a large number of hadronic events, but retain most of the gamma rays. The validity of this method of rejecting hadrons can be demonstrated convincingly by plotting the H/W distributions across all grid points for simulated gamma-ray showers and comparing them to those of hadron showers. As one can see in Fig. 6.10, gamma rays give rise to rather peaked distributions which show a clear maximum (the shower core) and a high degree of axial symmetry, whereas those from

[7] Trials with code using more than four background points, forming an actual ring around the $(H/W)_{max}$ position, did not reveal a measurable improvement of the method's accuracy.

[8] It is of some significance that the background points are picked within a plane orthogonal to the shower axis, and not in the horizontal plane like the points used for the grid alignment. This is necessary in order to avoid a strong elevation dependence of the H/W values at the background points. Since the offset points only serve as a means of quantifying the background, it does not matter that the chosen perpendicular plane has no significance whatsoever as far as shower physics are concerned.

hadronic showers are flatter and often irregular. The less peaked behaviour of the latter results in ξ values that are generally much higher than for gamma rays.

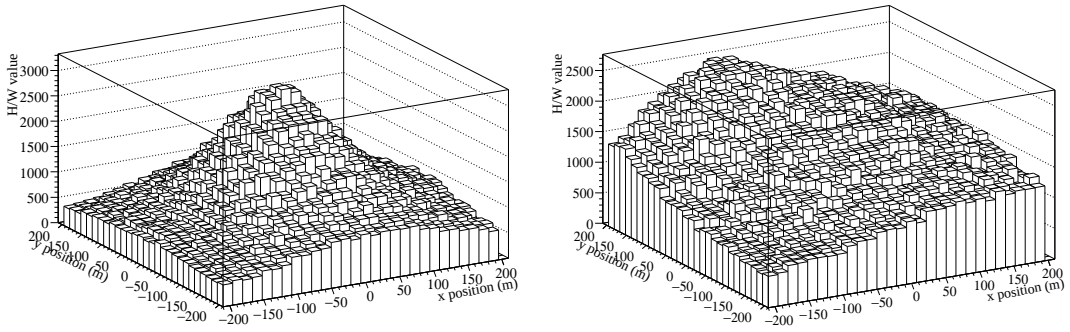

Figure 6.10: Surface plots of calculated height-to-width ratios for each grid point, for simulated data. Shown are the distributions due to a gamma-ray event (*left*) and from a typical proton-induced shower (*right*). The high degree of sphericity of the gamma-ray wavefront is reflected in the prominent peak and axial symmetry, giving rise to a large gradient between the H/W values at the peak (corresponding to the core location) and on the edges of the sampled area. From Lindner *et al.* (2007).

There are some technical points regarding the grid-alignment/grid-ratio technique which need to be mentioned here, particularly since it was used in the analysis described in this thesis. One concern is the exact choice of the distance d to the background points. Studies which used different values of d have shown that it should be larger than the maximum extension of the heliostat array, but not exceedingly large. Therefore, $d_0 = 200\,\mathrm{m}$ was chosen for a source at STACEE's point of highest sensitivity, the Crab transit point. However, the wavefront's degree of curvature decreases with decreasing source elevation, which is a result of the increasing distance to shower maximum (caused by a combination of the obvious geometrical effect and the fact that due to the increasing atmospheric overburden, shower maximum occurs higher in the atmosphere for more inclined air showers). If the quantity ξ is to be (more or less) independent of the actual pointing direction, the chosen value of d must take into account the different curvature scales. This is done by simply scaling d_0 with the distance to the expected shower maximum R, *i.e.* one sets

$$d = \frac{R}{R_{\mathrm{CT}}}\, d_0 \tag{6.6}$$

where $R_{\mathrm{CT}} = 11282\,\mathrm{m}$ is the predicted distance to shower maximum for the Crab transit point.

Another factor that greatly influences the hadron rejection power of the grid algorithm is the treatment of the closely spaced heliostats of the South cameras (clusters

6 and 7). Studies have shown that the value of ξ that maximises the quality factor depends sensitively on the core location (Lindner, 2006). Of particular concern is therefore the "pull" exerted by the channels of these two clusters on the trigger probability, which is due to higher mirror density on the ground as well as better optical throughput, and causes a preferential triggering of Cherenkov showers with cores to the south of the heliostat array centre. According to simulations, events with an above-average fraction of charge in the channels belonging to the South cameras require systematically higher ξ-cut values. Although this effect can probably be reduced to the more general core dependence of ξ that has been observed, a correction factor specifically addressing the relative charge of the South camera channels was applied to ξ. This factor differs from unity only for event where the charge in the two South clusters exceeds 30% of the total charge of the event, and it decreases linearly with increasing charge fraction. Even though this correction is entirely empirical, and it is obviously not straightforward to justify in a quantitative way why the two South camera clusters should be treated differently than the other clusters, including the correction factor consistently results in an improved hadron rejection power of the ξ-cut.

Finally, as already alluded to above, the fact that the delays are calculated for only one shower maximum height gives rise to energy selection effects (see also Sec. 3.1.1 for how the height of shower maximum depends on primary energy). Along with the less spherical shape of Cherenkov wavefronts from gamma rays of higher energies, this results in a gradual degradation of core resolution and hadron rejection power for gamma rays above ~ 1 TeV. Since the bulk of STACEE's gamma-ray events come from lower energy showers (and after all, its aim is to achieve a low energy threshold), optimising the discrimination capability of the method for gamma rays in the 10^{11} eV range is not entirely unwarranted.

Validation of the grid technique

In the preceding it was explained how the grid alignment technique works and how inherent biases are corrected for. All this by itself however is of little use if it is not backed up by evidence that the technique performs as advertised. Also, we need to address the performance of the technique and how the exact cut value on ξ is determined that will maximise the quality factor, as well as the cut value that is to be used in the actual data analysis.[9] There are two ways of validating the technique: applying it to simulated data, and to observations of the Crab Nebula. For various reasons,

[9]These two values of ξ are usually not the same. Knowing that the technique is sensitive to a large number of factors, that real data was taken in varying conditions and for a range of detector headings,

which include artifacts in the measured data related to the detector electronics which are not entirely understood and hence not modelled in the simulations, the latter is undoubtedly more assertive. However, since the technique was investigated and fine tuned via simulations, an agreement between Monte-Carlo predictions and results for the Crab Nebula would inspire confidence in the fact that all relevant processes are modelled satisfactorily by the simulations.

The grid-ratio cut has been tested on STACEE's observations of the Crab Nebula from 2002 to 2004. Applying a cut at $\xi < 0.35$ resulted in an enhancement of the raw (post-padding) signal from $4.8\,\sigma$ to $8.1\,\sigma$ (Lindner *et al.*, 2007). The right panel in Fig. 6.11 shows the ξ-distribution of the (ON−OFF) difference, superimposed on results of gamma-ray simulations. Despite the large error bars on the subtracted Crab data points, the distributions appear to match. Plots of Q versus ξ-cut value (see *e.g.* Lindner, 2006) suggest that a grid value cut at $\xi < 0.35$ yields the highest quality factor, and hence the highest final significance. Here it is remarkable that simulations and Crab data results agree very well in regard to the optimum ξ-cut value. The maximum achievable quality factor is predicted to be between 2.25 and 2.6, however in that case

and that cutting at too low a value of ξ is a sure way of losing most of the gamma-ray events, it is usually prudent to be somewhat conservative with the choice of ξ.

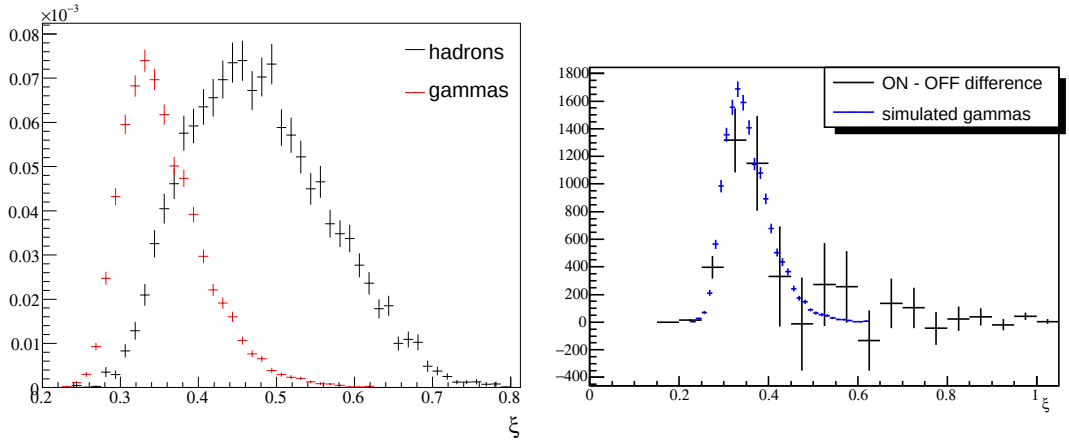

Figure 6.11: Number of events as a function of calculated ξ value. *Left*: Results for simulated gamma rays (*red*) and protons (*black*), for showers from the Crab transit point. The relatively good separation of the two distributions is encouraging. *Right*: Difference in the ξ-distributions of on-source and off-source events for the Crab data set of 2002-04 (ON−OFF, *black*), superimposed on the distribution for gamma-ray simulations (*blue*). The width of the gamma-ray excess distribution matches that of the simulations well, although statistics are low. From Lindner (2006).

only about 40% of all gamma-ray events pass the cut. In some cases, achieving a high detection significance is less important than keeping a large enough number of events; a slightly higher cut value may then be more beneficial. Since the Q-value falls off steeply as soon as the ξ-cut is reduced below that of the optimum yield, a less aggressive, *i.e.* more conservative cut may also be preferable for data taken at lower average elevations, or source coordinates significantly less ideal than those of the Crab Nebula.

CHAPTER 7

OBSERVATIONS OF H 1426+428

Having presented the basic analysis techniques for STACEE data on one hand and available simulation methods on the other, we shall now apply these two sets of tools to analysing STACEE's observations of the BL Lac type blazar H 1426+428, a well-known emitter of TeV gamma rays that was introduced in Sec. 2.6. We will begin with a description of the data sets and of the manual and automatic data selection cuts performed. The effects of processing the data sets by means of the procedures described in the preceding chapter are examined and the final results presented. Next we will describe the results of detector simulations customised to the data at hand, and how they are combined with the output of the data analysis pipeline in order to yield a result of scientific value.

The last part of this chapter is devoted to a discussion of the obtained results. A comparison with previous flux measurements in the gamma-ray energy range is made, and we will also take a brief look at the RXTE-ASM count rates coincident with STACEE's observations. In the end, possible implications of our result are discussed in regard to the SED of the extragalactic background light, as well as to primary and secondary gamma-ray emission from H 1426+428.

7.1 Data sets and basic analysis

The right ascension of H 1426+428 (14 h 28') allows its observation by ground-based gamma-ray telescopes from February through July. The source was observed with STACEE during the observing periods of three consecutive years, from 2002 to 2004. For various reasons, the data taken during the spring of 2002 are of limited value: the detector had just been upgraded from 48 to 64 heliostats and was not yet operating in a stable manner, important calibration constants such as the PMT gains were not monitored at regular intervals, and only 32 of 64 waveform digitisers were available. For these reasons, the data set of 2002 is not considered here. The subjects of this analysis are the data sets of 2003 and 2004, which are of higher quality and also of larger size.

There are two fundamental differences between the two data sets that concern us here. Firstly, the heliostats were canted normally (*i.e.* monocanted) during the 2003 observations, and according to the paracanting scheme in 2004. As mentioned previously, paracanting facilitates event reconstruction particularly for low-energy events, but results in a slight increase of the energy threshold (Sec. 6.6.1). We do not perform event reconstruction for any of the data considered; the expected flux from H 1426+428 is too low for a spectral measurement to be possible, and we chose the grid alignment technique for hadron rejection over the core/direction fit. In the framework of this analysis we hence do not benefit from the advantages of paracanting.

The second difference concerns the strategy according to which the discriminator thresholds were chosen. In order to maximise the detector's sensitivity to low-energy events, the operating thresholds were generally not static, but chosen by means of rate-versus-threshold measurements (see Sec. 4.5.4) each observing night just before observations of H 1426+428 commenced. For the 2003 data, the thresholds were chosen with the aim of facilitating offline analysis, and depending on the breakpoints of the rate-versus-threshold curves the observer decided between only two threshold values. In the 2004 observing season, the choice was made to set the thresholds to a fixed amount (usually 10 - 15 mV) above the breakpoints. This more "aggressive" way of choosing the operating points in the later season partially made up for the loss of effective area due to paracanted heliostats. The nominal thresholds were 140 mV for most pairs of 2003's data set, with the exception of four pairs that were taken at 120 mV. For the data of 2004, the set thresholds ranged from 110 to 130 mV. The difference in average thresholds is not entirely due to observation strategy, but can in part be attributed to different values of the PMT gains, which were on average higher in the earlier data set.

The 2003 observations of H 1426+428 were performed from March 9^{th} to June 29^{th} (UT). The data set comprises 28 regular (28 minute) and 2 short (14 minute) pairs, amounting to a total on-source time of 48.7 ksec distributed over 20 observing nights. In 2004, the source was observed between March 15^{th} and June 19^{th}. A total of 85 regular and 3 short pairs were taken during this time, corresponding to an observing time of 140.3 ksec during 37 observing nights. Unfortunately, due to a combination of weather factors and ambient man-made disturbances (*e.g.* RF interference), the runs of the latter data set were on average less clean than the data of the previous year, and a larger fraction of it had to be rejected. In the following sections we will describe the effects of the basic analysis on the two data sets, including time cuts and software padding.

7.1.1 Data quality cuts

The first step in the selection of a clean and bias-free data sample was the rejection of obviously flawed pairs by means of a manual inspection of the corresponding L1-rate *vs.* time graphs, such as the ones shown in Fig. 6.5. Fluctuating L1-rates indicate clouds passing through the field of view; these are usually thin, high clouds that were not visible when the data were recorded. Sometimes entire observing nights were affected by poor sky conditions, and all runs taken during the night ended up being rejected in the offline analysis. Pairs for which only part of the overlapping time was affected by clouds were kept, and the compromised time intervals were subsequently excluded from further analysis by the L1-ratio algorithm (Sec. 6.4.3).[1]

Next, using all data pairs that passed visual inspection, automated time cuts were performed following the various procedures described in Sec. 6.4. For the L1-ratio algorithm a cut window of $3\,\sigma$ around the Gaussian peaks was chosen, as is shown in Figures 6.6 and 6.7. The on-source live times before and after the two stages of data selection cuts are listed in Table 7.1, including the percentages of live time remaining after all time cuts. Note that although the total observing time of H 1426+428 was almost three times longer in 2004 than in 2003, only half of the larger data set passes the time cuts, while about two thirds of 2003's data are accepted. The final hour-angle exposure distributions for both data sets, which include only those time intervals that passed all time cuts, are shown in Fig. 7.1. Only a small, yet still significant portion of both data sets corresponds to source elevations of less than 60° ($HA \gtrsim 40°$).

Data set	observing time	pair selection	quality cuts	fraction remaining
2003	48 720 s	40 320 s	31 977 s	65.6 %
2004	140 280 s	108 360 s	70 790 s	50.5 %

Table 7.1: On-source observing times (in seconds) for both H 1426+428 data sets before and after the two stages of time cuts, *i.e. pair selection* (or rejection) by manual inspection of L1-rates and automated data *quality cuts*. Also shown are the fractions of the original observing times that are left after all time cuts.

[1] As part of a quick analysis, one may opt to feed *all* pairs to the L1-ratio algorithm and let it flag the flawed pairs for removal. For a more thorough analysis this is not recommended, as long sections of wildly fluctuating L1-rates in ON and/or OFF runs may skew the L1-ratio distributions upon which the final L1-rate cuts are based.

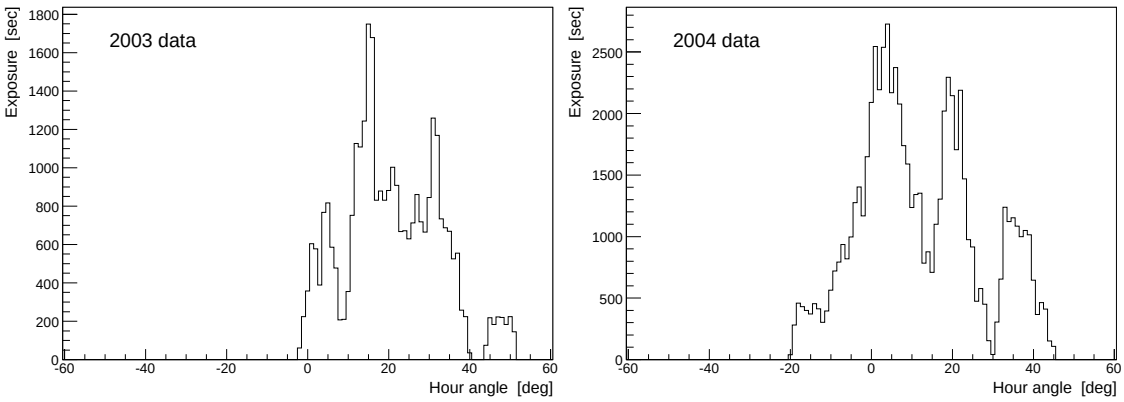

Figure 7.1: Hour-angle exposure distributions in units of live-time seconds after time cuts for the 2003 (*left*) and 2004 (*right*) observations of H 1426+428.

7.1.2 Promotion corrections

Before the two halves of a *cleaned* ON/OFF pair, consisting of sections of time that passed the aforementioned quality cuts, can be compared to each other, the variances of the FADC traces of both halves first need to be equalised. Only then can the pairs be used to calculate a significance. As described in Sec. 6.3.2, this was done by adding sample traces of an appropriate variance to the traces of the less noisy half-pair. For our two data sets, the traces with the lower variances were usually those of the OFF runs. The upper panel of Fig. 7.2 shows the average (ON − OFF) PMT anode current differences for the 2003 data set on a run-by-run basis. As we can see, the current differences are mostly positive, but vary significantly; for some pairs they are even negative, meaning the OFF run is brighter than the ON run. In all cases, however, they are very small, scattered around a mean of $+0.2\ \mu$A and not exceeding $+0.7\ \mu$A. The data set of 2004 is distributed very similarly, with the exception that it contains several pairs during which the currents of all channels suddenly dropped by $\sim 1\,\mu$A at one point in either the ON or OFF run. This is most likely due to a bright light being switched off somewhere outside of the area of the NSTTF, increasing the amount of software padding required in the affected pairs. Fig. 7.2 also shows that there appears to be no correlation between raw, pre-padding significances (*i.e.* ON-excess rates) of individual pairs and their average PMT currents or PMT rates.

In order to determine the amount by which the discriminator thresholds should be raised in the offline trigger reimposition, we examine Fig. 6.3. We note that the average (ON − OFF) current difference of the dimmest of the observed stars, HIP 89279, is $\Delta I = (1.80 \pm 0.04)\,\mu$A, and thus by far larger than the current differences in the

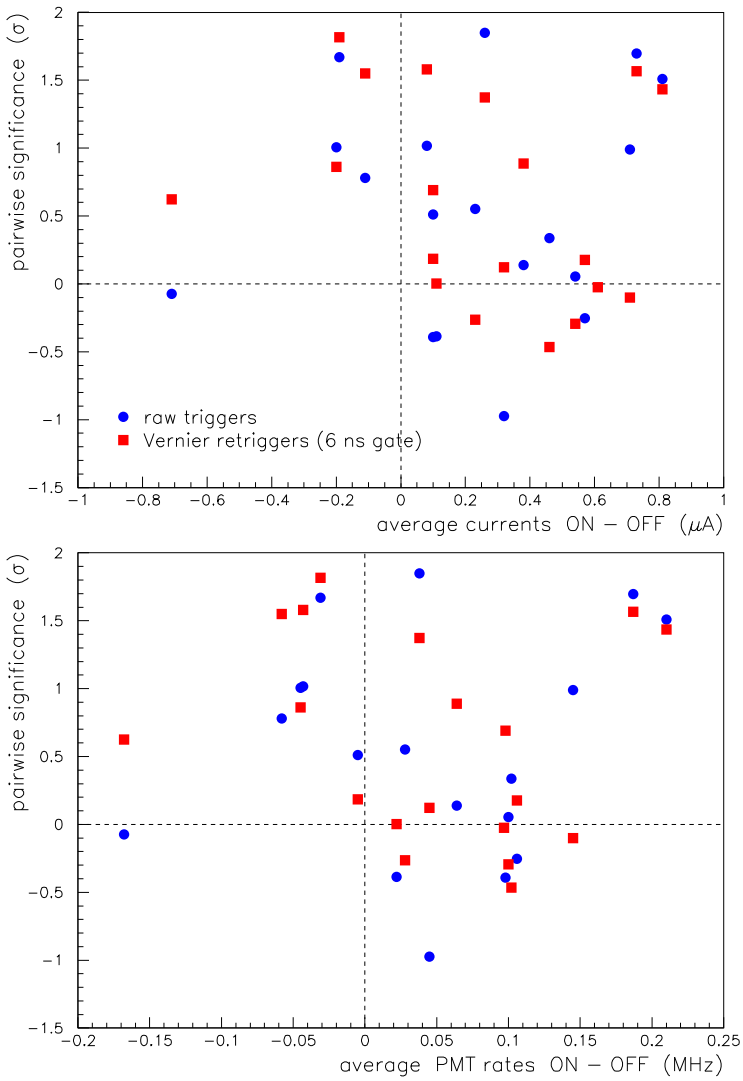

Figure 7.2: Average differences in PMT anode currents (*top*) and PMT rates (*bottom*) between ON and OFF run for individual pairs of the H 1426+428 (2003) data set. Plotted on the y-axis are pair-wise significances: raw data (*blue circles*) and after trigger reimposition with a static, narrow gate of 6 ns width, but without padding (*red squares*).

H 1426+428 data sets. The final, padded excess rate of this star as a function of threshold raise is plotted in Fig. 7.3. It appears that a threshold increase of $X = (15 \pm 1)$ mV suffices to reduce the promotion excess to zero.[2]

Raising the discriminator threshold is equivalent to increasing the energy threshold and degrading the sensitivity of the measurement. The degree to which the sensitivity

[2]Curiously, the value of X that causes the excess to vanish seems to be similar for observations of the star *Hercules 59*, which gave rise to much larger current differences (3.7 μA; see Fig. 6.4). This apparent agreement is probably a result of the large uncertainties affecting both measurements; increasing (ON−OFF) current differences generally do require larger values of X (Lindner, 2005).

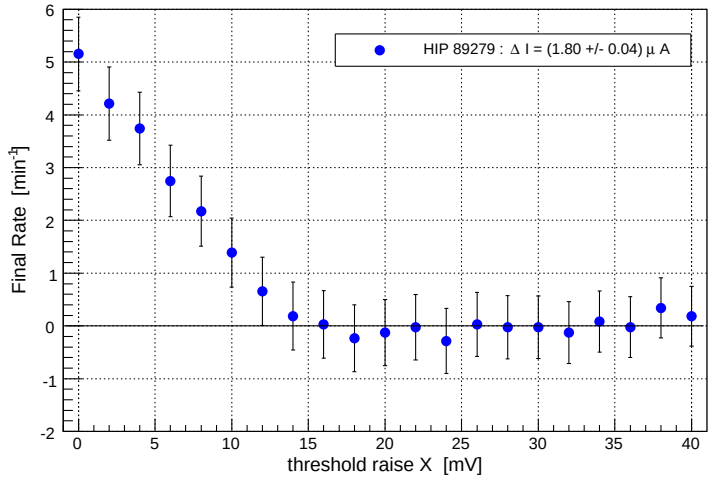

Figure 7.3: Final excess rate after all time cuts for observations of the star *HIP 89279* ($m_B = 5.98$) in the summer of 2003, as a function of threshold increase X. Heliostats were monocanted. The average current difference for this star was $\Delta I = (1.80 \pm 0.04)\,\mu$A, the raw excess rate was (8 ± 2) min^{-1} (Lindner, 2006b). Note that the calculated rates are correlated, hence the jitter in the points is less than if it were due to independent measurements (*i.e.* as expected from the error bars).

is reduced depends on the steepness of the gamma-ray emission spectrum of the source. Since the gamma-ray spectrum of H 1426+428 is steep ($\alpha \simeq 3.5$) at energies just above STACEE's sensitivity range, and the expected fluxes are low, we chose to increase the thresholds by a small amount only. For the data set of 2003, the thresholds were raised by 8 mV above the calculated effective thresholds. For the 2004 data set, the thresholds were raised by 12 mV. Given the small values of ΔI for our data sets, the above values of X are expected to allow for the padding to work, yet they do not result in an excessive increase of the energy threshold. Due to the sudden current drops seen in part of 2004's data set (see above), the threshold increase was chosen to be slightly higher for these data than for the pairs of 2003.

During the trigger reimposition in `Pass0`, the standard gate width of 16 ns was used for all data. The significances and excess rates of the two data sets resulting from the basic ON-OFF analysis – before as well as after the promotion correction – are given in Table 7.2. Since STACEE's sensitivity is low without the implementation of any kind of hadron rejection algorithm, we shall leave a discussion of these intermediate results for later.

H 1426+428: 2003 observations (25 pairs)							
Cuts	$t_{\mathrm{ON}}\,[s]$	$t_{\mathrm{OFF}}\,[s]$	N_{ON}	N_{OFF}	ΔN	$S\,[\sigma]$	$R\,[\mathrm{min}^{-1}]$
none (raw)	31977.30	32012.32	158536	157486	1050	2.18	2.29 ± 1.05
padding	31977.30	32012.32	133495	133063	432	1.12	1.08 ± 0.97
padding + grid	31977.30	32012.32	24480	24162	318	1.56	0.65 ± 0.41

H 1426+428: 2004 observations (68 pairs)							
Cuts	$t_{\mathrm{ON}}\,[s]$	$t_{\mathrm{OFF}}\,[s]$	N_{ON}	N_{OFF}	ΔN	$S\,[\sigma]$	$R\,[\mathrm{min}^{-1}]$
none (raw)	70789.96	70845.56	376173	375361	812	1.28	0.94 ± 0.73
padding	70789.96	70845.56	281704	281251	453	0.90	0.57 ± 0.64
padding + grid	70789.96	70845.56	49093	49065	28	0.21	0.06 ± 0.27

Table 7.2: Final ON-OFF statistics for the complete H 1426+428 observations of 2003 (*upper table*) and 2004 (*lower table*): after time cuts only (*top*), after software padding and static gate cut (*centre*) and after an additional grid-ratio cut with $\xi < 0.37$ including the correction for the South camera fraction (*bottom*). t_{ON} and t_{OFF} are the total on- and off-source live times, respectively; N_{ON} and N_{OFF} are the on- and off-source event numbers passing the cuts, ΔN is the number of on-source excess events, and S the significance. The excess rates R are quoted in units of counts per minute.

7.2 Gamma-hadron separation

As the reader may recall from Sec. 6.6, the final stage in the data analysis is the application of event cuts with the aim of rejecting a large fraction of the hadronic background while retaining as many gamma-ray events as possible. For this analysis, we chose to apply the grid-ratio cut (Sec. 6.6.2), because it is proven to perform better than any of the other gamma-hadron separation techniques. A promising option to be explored in this context is the possibility of applying other types of event cuts in conjunction with the grid-ratio cut. For various reasons however this did not lead to improvements, but only to an (unwanted) reduction of the event rates:

- It was found that the combined core and direction fit by means of charge templates (Sec. 6.6.1) is highly correlated with the grid-alignment technique, and hence does not represent an independent hadron-rejection method. This is not surprising, as both methods fit for the shower core and exploit the sphericity of the Cherenkov wavefront. Among the two techniques, the grid-alignment method was found to yield higher quality factors.

- Experimenting with preferentially selecting low-energy gamma-ray showers by rejecting all events with one or more saturating channels in the FADC traces (*i.e.* with pulse amplitudes $> 1\,\mathrm{V}$) did not increase the significance of Crab Nebula

observations, but reduced the final excess rate. Using these types of cuts on the H 1426+428 data sets revealed the same behaviour.

- Other existing hadron-rejection techniques do not yield promising quality factors.

7.2.1 Grid-ratio cut for H 1426+428 data

When applying the grid-alignment/grid-ratio cut, it was decided to include the empirical correction for the South camera fraction as described in Sec. 6.6.2, since it leads to verifiable improvements of the achievable quality factors.[3] The next question needing to be addressed is at which exact value of ξ the grid-ratio cut should be applied. As already mentioned in the last chapter, a cut value of around 0.35 appears to maximise the quality factor in simulations and in data from observations of the Crab Nebula (Lindner, 2006,b). For our two data sets on H 1426+428 however, this cut value may not be appropriate. It is known that the distributions of ξ-values for simulated gamma-ray as well as hadron spectra shift towards higher values of ξ as the source elevation decreases (Lindner, 2006), which is partially a result of the increasing energy threshold and partially due to changes in the geometry. A significant fraction of the data on H 1426+428 was recorded at hour angles (HA) larger than 30°, *i.e.* at elevations lower than 65.5°. If we chose to optimise our analysis for low-energy gamma-rays, we could implement an hour-angle cut of $HA < 35°$ combined with a grid-ratio cut $\xi < 0.35$. This however is not the aim of this analysis, because as we shall see in the next section, STACEE did not detect a signal from H 1426+428. For detectors sensitive only to the integral flux above some energy threshold, upper limits become more stringent the more time is spent observing the source, independent of energy threshold. Since the runs taken at larger zenith angles appear to be rather clean, we see no reason not to include them. For this reason, $\xi < 0.37$ was chosen as a cut value for the grid alignment.

7.2.2 Detection significances

For each individual ON/OFF pair as well as for the two data sets as a whole, the significances S after software padding and after the grid-ratio cut were calculated according to Li & Ma (1983). A pair-by-pair listing of raw (pre-padding) and final (*i.e.* after padding and grid cut) significances for both data sets can be found in Appendix B. The

[3] As described in Sec. 6.6.2, the relevant studies of the South-camera correction were made using simulations as well as Crab Nebula observations. For both H 1426+428 data sets, the correction was also found to increase the final significance for any tested cut value of ξ. This fact on its own does of course in no way influence our choice of cuts, but it is interesting nonetheless.

combined results for the two observations are shown in Table 7.2. Both data sets have a positive raw significance, probably due to the – on average – slightly higher background levels in the on-source halves of the pairs. In both cases, the raw excess is reduced by the padding cut. While the grid cut further reduces the significance of the 2004 data to essentially zero, this does not happen for the data taken in 2003. The small remaining excess ($1.56\,\sigma$) in this data set may be indicative of a real gamma-ray signal, although it is of course compatible with zero signal.

The live-time weighted significance distributions of individual pairs are plotted in Fig. 7.4. The distribution for the data of 2004 is centered at zero and looks reasonably compact and symmetric, which is what one would expect in the case where no significant source of gamma rays is in the detector's field of view. For 2003's data, some degree of symmetry is recognisable as well, but the distribution is more spread out and clearly shifted towards positive significances. Since the TeV gamma-ray flux from H 1426+428 has been seen to vary significantly between the observing seasons of subsequent years (recall Sec. 2.6.2), which is a blazar-typical behaviour, we shall not combine the two data sets to produce a single significance and excess rate. Merging the two data sets would only be useful if it resulted in a high enough significance for a detection to be probable, regardless of the respective two flux states of the source; this however is not the case here, as even the combined data set does not yield a significant number of excess events.

7.3 Detector Response

In order to convert the results of the ON-OFF data analysis that were just presented to scientifically meaningful quantities such as absolute gamma-ray fluxes or upper flux limits, they need to be calibrated by means of appropriate detector simulations. This is done by first determining the *effective area* of the detector as a function of primary gamma-ray energy. The effective area for a particular primary energy essentially corresponds to the virtual area, in a plane perpendicular to the direction of the incident gamma rays, that the detector would have if it were able to record (*i.e.* trigger on) every single gamma-ray event of that energy. In gamma-ray astronomy, effective area is usually used to quantify the detector response to a gamma-ray point source, *i.e.* it is associated with a particular source direction, and hence carries the unit of area (m^2). For a complex detector such a STACEE, the effective area is a function not only of energy, but also of the detector heading, the employed heliostat canting scheme and a multitude of electronic parameters (PMT gains and thresholds, PMT currents *etc.*; see Sec. 5.4).

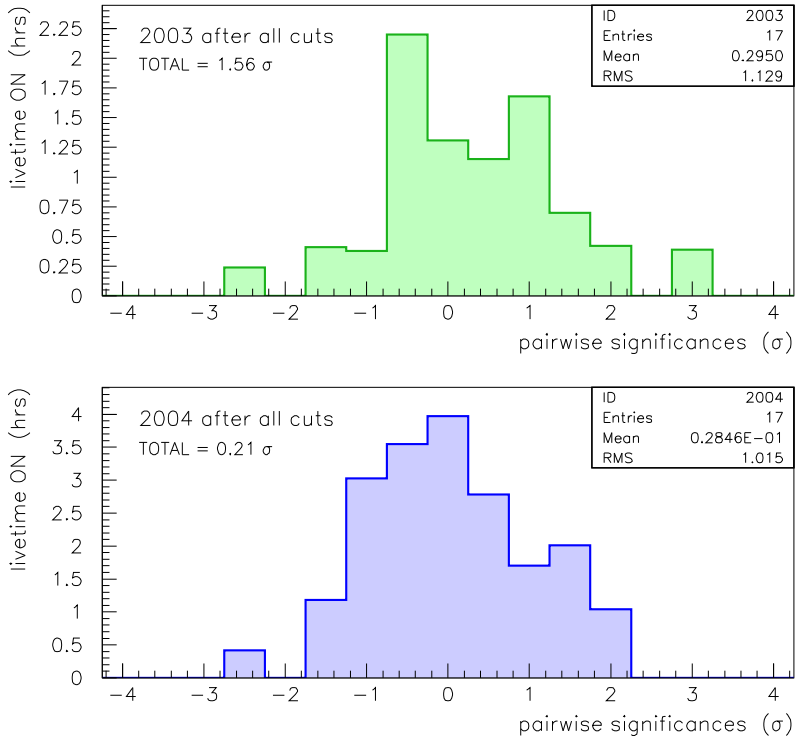

Figure 7.4: Exposure-weighted pair-wise significances of 2003 (*top*) and 2004 (*bottom*) observations of H 1426+428. The y-axis shows the accumulated live time on source per σ-bin, in units of hours.

Once the appropriate effective area function $A_{eff}(E)$ is known, detected on-source excesses (or the lack of an excess) can in principle be converted to corresponding gamma-ray fluxes. If $(\mathrm{d}N/\mathrm{d}E)(E)$ is the differential gamma-ray flux from the observed point source in units of $[\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{TeV}^{-1}]$, the relationship between effective area and observed gamma-ray rate R_γ (in units of Hz) is

$$R_\gamma = \int_0^\infty A_{eff}(E)\,\frac{\mathrm{d}N}{\mathrm{d}E}(E)\,\mathrm{d}E \tag{7.1}$$

If it were possible to measure the energy E of each recorded event, the above integral could be written as a sum of integrals over different energy intervals with means E_i, and the measured rates $R_{\gamma,i}$ for each energy bin could be used to reconstruct the initial gamma-ray spectrum. The fact that very few gamma-ray sources are strong enough for

STACEE to perform spectral measurements[4], however, means that we will only be able to make a statement in regard to the *integral* gamma-ray flux from H 1426+428 above some energy threshold, and not about the actual shape of the spectrum. Also, for power-law spectra, knowing the integral flux is not enough to resolve the degeneracy between spectral normalisation and power-law index. In order to compare our measurement with the actual SED – and last but not least with differential flux measurements of other detectors – we first need to assume a certain value of the spectral index α of the source spectrum in the energy interval to which STACEE is sensitive. We shall get back to this step in Sec. 7.3.2. The following sections present details on how the simulation results necessary for building up a set of effective areas are obtained, and how our rate measurements can be calibrated with the latter.

7.3.1 Effective Area for gamma rays

Since the detector's effective area depends on a rather large number of variables, obtaining an appropriate, averaged effective area function for a given set of observations clearly requires detector simulations tailored to these observations. Several choices are to be made as to the desired energy and angular resolution of the effective area values. If the step size is too small, the time required for running the simulations may be too long; if the energy or angular resolution of the simulations is too coarse, the errors on the result will be too large.

Energy binning

For obtaining the desired energy resolution, large sets of gamma-ray showers at discrete energy steps were produced. The energies of the generated showers started at 20 GeV and increased in steps of 10 GeV up to 180 GeV; above that energy, the energy step size was roughly logarithmic.[5] With the purpose of allowing the simulations to run more efficiently, the scattering radius $R_{\rm scat}$ used in `Sandfield` (Sec. 5.3) was adapted to the primary energy, ranging from 150 m for 20 GeV to 350 m for 5 TeV. The total number of simulated showers decreased logarithmically with energy, starting from 10 000 CORSIKA showers at 20 GeV to 200 at 5 TeV. Each CORSIKA shower was processed 10 times by `Sandfield`, hence the aforementioned numbers need to be multiplied by 10 in order to obtain the actual number of showers used for each energy.

[4] Only gamma-ray fluxes from the Crab Nebula and Mrk 421 are high enough for spectral reconstruction to be possible with STACEE.

[5] The primary energies of these sets of gamma-ray showers were 200, 250, 300, 400, 600, 1000, 2000 and 5000 GeV.

Detector headings

As the position of the source in the sky varied greatly in both of our data sets, effective area curves (*i.e.* graphs of effective area as a function of gamma-ray energy) needed to be obtained for a range of different detector headings. The arrival directions of gamma-ray air showers – which, since we simulate a gamma-ray point source and thus do not scatter the primary directions in solid angle as for cosmic rays, are identical to the corresponding detector headings – were chosen at various points along the track of H 1426+428 through the sky, whose local sky coordinates (azimuth and elevation) are determined by the geographic latitude of the observatory (34.96°) and the declination of H 1426+428 ($\delta = 42.67°$). The simulated detector headings included the *transit point* (*i.e.* the point of zero hour angle: azimuth = 0.0°, elevation = 82.29°) and points spaced at intervals of 10° in hour angle to either side of transit up to $HA = \pm 50°$. The exact coordinates of the simulated directions are listed in Table 7.3. Since the canting scheme was different for the two data sets at hand (monocanting in 2003 and paracanting in 2004), two full sets of effective area curves were generated, one each with monocanted and paracanted heliostats.

Simulated detector headings			
HA	δ	azimuth	elevation
−50 °	42.67 °	63.19 °	50.87 °
−40 °	42.67 °	63.79 °	58.21 °
−30 °	42.67 °	62.60 °	65.54 °
−20 °	42.67 °	57.61 °	72.67 °
−10 °	42.67 °	42.26 °	79.05 °
0 °	42.67 °	0.00 °	82.29 °
+10 °	42.67 °	317.74 °	79.05 °
+20 °	42.67 °	302.39 °	72.67 °
+30 °	42.67 °	297.40 °	65.54 °
+40 °	42.67 °	296.21 °	58.21 °
+50 °	42.67 °	296.81 °	50.87 °

Table 7.3: List of detector headings simulated along the track of H 1426+428. Headings are given in equatorial coordinates (hour angle *HA* and declination δ), and in local horizon coordinates (azimuth and elevation).

Emulation of real on-source runs

Having two sets of `Sandfield` files for different gamma-ray energies and detector headings at our disposal, we still need to take into account the different environments in

which the observations were made. The term "environments" in this case refers to all the variable input parameters required by the final stage in the simulation chain, the electronics simulation. The three main sets of parameters that need to be modelled properly for each pair, and the causes of their variation, are:

- The discriminator thresholds were not static, but were set according to the breakpoints of the rate-versus-threshold curve measured immediately before the observations. The breakpoints depended on absolute sky clarity and on the source elevation. Setting the discriminator thresholds correctly in the simulation is important, as the energy response of the detector depends sensitively on the actual thresholds values.

- The PMT anode currents systematically increased with increasing zenith angle, owing to higher night-sky background levels (scattering in the atmosphere) at lower elevations.

- The PMT gains showed a slight decrease over time. As both observations extended over several months, the gain variation within each data set is significant.

Since the "binning" of the simulated data is determined by the `Sandfield` files for the different detector headings, there are at least as many `Elec` parameter files available to describe the whole data set as there are simulated source directions. Appropriate values of the above parameters were thus obtained by running a customised program on the real data that filled two-dimensional histograms with all relevant quantities. In a second pass, the program binned these quantities according to hour angle corresponding to the simulated detector headings. A set of parameter files for the electronics simulation was then written out for each detector heading, containing averaged values of the parameters for each *HA*-bin. Finally, the electronics simulation was run on the `Sandfield` shower files for each detector heading using the proper parameter file. From the text output files, the trigger probabilities $P_i(HA, E)$ for each of the simulated energies E, each hour-angle bin HA and both data sets (i) were extracted by means of a `perl` script. From these, the raw effective areas $A_{eff,i}$ can easily be calculated via

$$A_{eff,i}(HA, E) \;=\; \pi\, R^2_{\mathrm{scat}}(E)\; P_i(HA, E) \tag{7.2}$$

where $R_{\mathrm{scat}}(E)$ is the scattering radius used when generating the `Sandfield` shower files of the energy E. Those input parameters which were not subject to change between individual runs of the electronics simulation can be found in the sample parameter file listed in Appendix A.3.

After repeating this process for all detector headings, we end up with a set of raw effective area curves for each data set i, consisting of curves for a number of different detector headings (HA). These are plotted in Fig. 7.5. Upon inspecting the curves one may notice that the sensitivity to low-energy gamma rays degrades rapidly as the source departs from the transit point, which is a result of the lower collection efficiency of the STACEE optics as well as the decrease in Cherenkov light density on the ground. At very high energies ($E \gtrsim 600\,\mathrm{GeV}$), on the other hand, the effective area increases with increasing zenith angle. This is caused by the larger cross-section of the Cherenkov light emitting volume of high-energy showers for more inclined arrival directions, allowing the detector to sample showers from primaries with increasingly large impact parameters.

Cut efficiencies

The sets of effective areas presented above describe the *raw* sensitivity of the detector, *i.e.* the sensitivity at the trigger level. These effective areas can hence be compared to the raw on-source excess, if there were no systematic differences in the on- and off-source runs (such as different field brightnesses), but without applying a padding cut (which increases the energy threshold). In order to obtain effective areas that correspond to the final results of the analysis, the simulated data must be treated in exactly the same way as the real data. We must therefore apply the same event cuts to both real and simulated data.

To this end, the fake data (`stoff` format) output files of the electronics simulation that were generated at the same time as the text output used for the effective area calculation were processed with the basic and advanced analysis programs `Pass0` and `Pass1` (see Chapter 6). The simulated data consist only of gamma-ray showers, hence they are treated in the same way as the real on-source runs. In order to keep the procedural effort on a justifiable level, we make the assumption that the variances of the on-source runs are *always* higher than those of the corresponding off-source runs. In reality this is not the case for every pair of the two present data sets, however in all cases the variance differences are very small, much smaller than the uncertainties in the relevant simulation parameters (PMT thresholds, gains *etc.*). One can therefore safely ignore these cases of negative variance differences, and no software padding is applied to the simulated data. The padding cut is emulated by reimposing the trigger conditions at the appropriate (higher) thresholds, and the grid-ratio cut is applied with the same

cut value (0.37) as in the real data.[6] In this way a *cut efficiency* is calculated for each simulated gamma-ray energy, equalling the fraction of triggered events that pass the combination of padding and grid-ratio cuts. The raw effective areas are then multiplied by the cut efficiency, yielding our final, or *net* effective areas.

Fig. 7.6 shows the net effective areas for the two H 1426+428 data sets. The most striking difference to the raw curves (Fig. 7.5) is the decline of the effective area with increasing gamma-ray energy after a broad peak at $\sim$170 - 400 GeV (depending on *HA*). This is caused by the grid-ratio cut, which is optimised to lower energies (recall Sec. 6.6.2). The second, more subtle difference is the slightly slower turn-on – a result of the trigger reimposition at a higher threshold, *i.e.* the padding cut.

7.3.2 Response Functions and Energy Thresholds

From a known set of effective area functions $A_{eff}(HA, E)$ and an on-source excess rate R_γ (presumably due to gamma rays), one may hope to determine the differential flux $(\mathrm{d}N/\mathrm{d}E)(E)$ of gamma rays from the observed source by applying Eq. 7.1. As mentioned previously, STACEE has no energy resolution for all but the strongest gamma-ray sources. In order to convert an observed rate R_γ to an integral flux or – equivalently – to be able to quote a differential flux measurement at one single energy, one must make some assumptions about the shape of the original gamma-ray spectrum. Over sufficiently small energy intervals it is usually correct to fit the gamma-ray spectrum with a simple power law (Eq. 2.31); see Sec. 2.3 for arguments why this is usually a valid approximation. If we thus assume the spectrum to have the form

$$\frac{\mathrm{d}N}{\mathrm{d}E}(E) \;=\; N_0 \left(\frac{E}{E_0}\right)^{-\alpha} \tag{7.3}$$

where N_0 is the spectral normalisation, carrying the units $[\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{TeV}^{-1}]$, and E_0 an arbitrary constant energy that determines the normalisation N_0 and which we choose to be $E_0 = 1$ TeV, we can rewrite Eq. 7.1 in the form

$$R_\gamma \;=\; N_0 \int_0^\infty A_{eff}(E) \left(\frac{E}{E_0}\right)^{-\alpha} \mathrm{d}E \tag{7.4}$$

The spectral index α can often be constrained by means of other flux measurements in the energy range, or even by assuming a certain gamma-ray emission model and level of

[6] As explained in Sec. 6.6.2, a comparison between simulations and Crab Nebula observations showed that the parameter ξ is not affected by any systematic differences between simulated and real data, hence a cut on ξ will have the same effect in both domains.

EBL-absorption. Using a given value of α, the only free parameter that determines the differential flux spectrum is N_0, which can then be calculated via

$$N_0 = \frac{R_\gamma}{\int_0^\infty A_{eff}(E)\,(\frac{E}{E_0})^{-\alpha}\,\mathrm{d}E} \tag{7.5}$$

The function in the integrand $(A_{eff}(E)\,(E/E_0)^{-\alpha})$ is also called the *response function*. The energy integral of the response function determines the detector's sensitivity (in dependence of α), and, for a given source flux, is proportional to the expected final gamma-ray rate.

Measurements of the differential flux from H 1426+428 right above STACEE's energy range indicate that $\alpha = 3.50 \pm 0.35$ (see Chapter 2 and Petry *et al.*, 2002).[7] In order to characterise STACEE's response to the gamma-ray flux from H 1426+428 in the $\sim$60 - 600 GeV energy range, we assume that the spectrum in this energy band is an extension of the SED at higher energies. We thus obtain the required response functions by folding the effective area curves with a power-law spectrum of index $\alpha = 3.5$. The resulting graphs are plotted in Fig. 7.7 for all relevant hour-angle values. As expected, the detector sensitivity (*i.e.* the integral of the response curve) decreases rapidly with increasing *HA*; the effect is enhanced by the steepness of the assumed spectrum. Evident is also the somewhat lower sensitivity for the paracanting scheme used during 2004's observations (lower panel) when compared to that for the observations in 2003.

A gamma-ray telescope's energy threshold is commonly defined as the peak of the response curve, which is of course dependent on the spectral shape. The threshold energy corresponds to the energy of maximum sensitivity and is hence used as a reference energy for quoting integral as well as differential fluxes. According to Fig. 7.7, if all observations had been performed within 10° of transit and with monocanted heliostats, STACEE would have a final energy threshold of 150 GeV.

Exposure-averaged response functions

While examining the detector response as a function of hour-angle is more informative, the calculation of a flux (or upper limit) from our observed rate requires the proper weighting of the response functions for individual *HA*-values, and the construction of an *exposure-averaged* response function. This averaged response function will also yield the actual energy threshold for the observations.

[7] The spectral measurement by CAT has larger errors and is a time-average over three consecutive years; the flux measurements by HEGRA lie at higher energies where the power-law spectrum is likely to be modified by EBL-absorption.

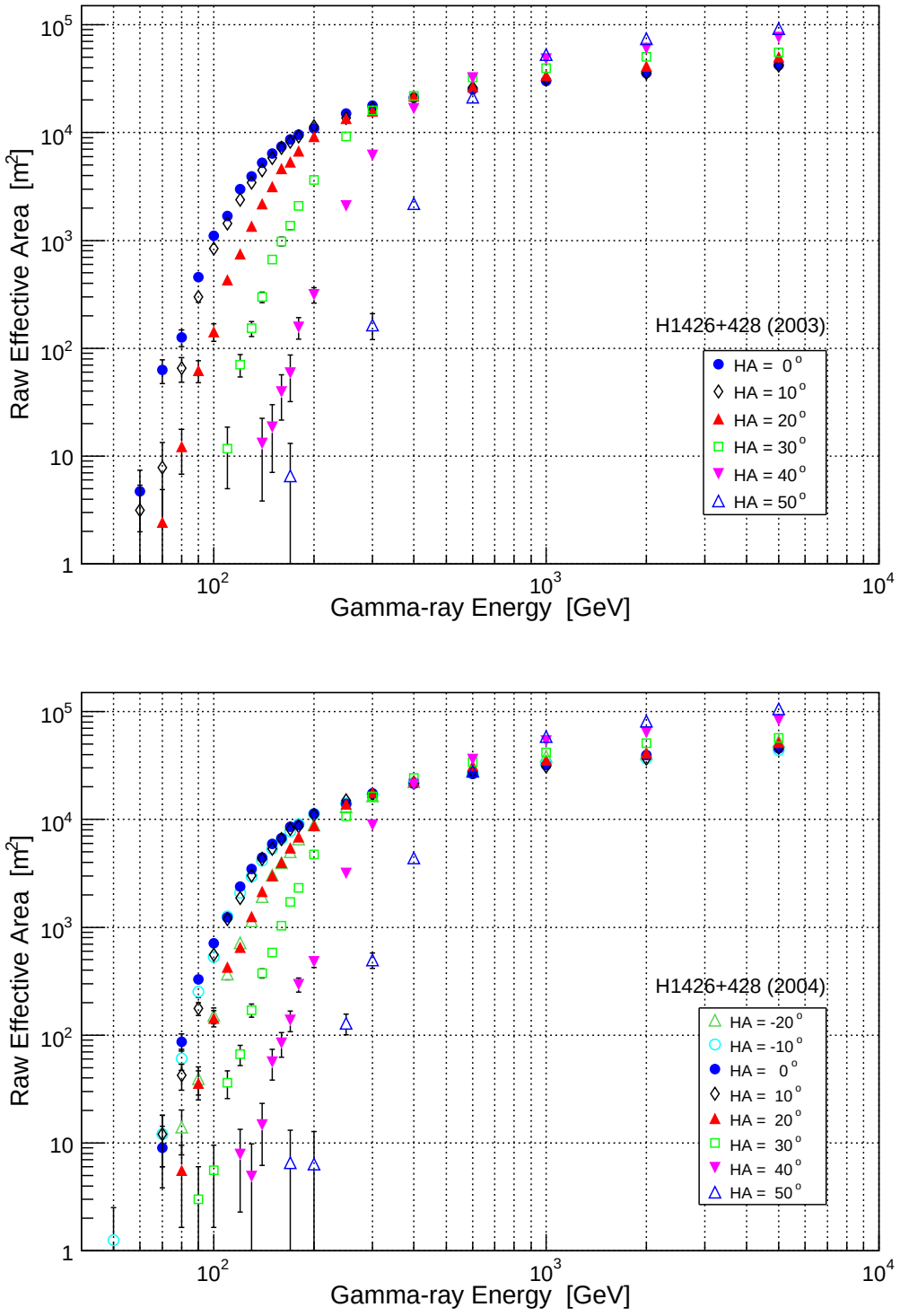

Figure 7.5: Raw effective areas (*i.e.* before the application of event cuts) for the 2003 (*top*) and 2004 (*bottom*) observations of H 1426+428, for all 10° hour-angle (*HA*) bins for which there were observing time intervals passing the time cuts. Statistical errors are plotted.

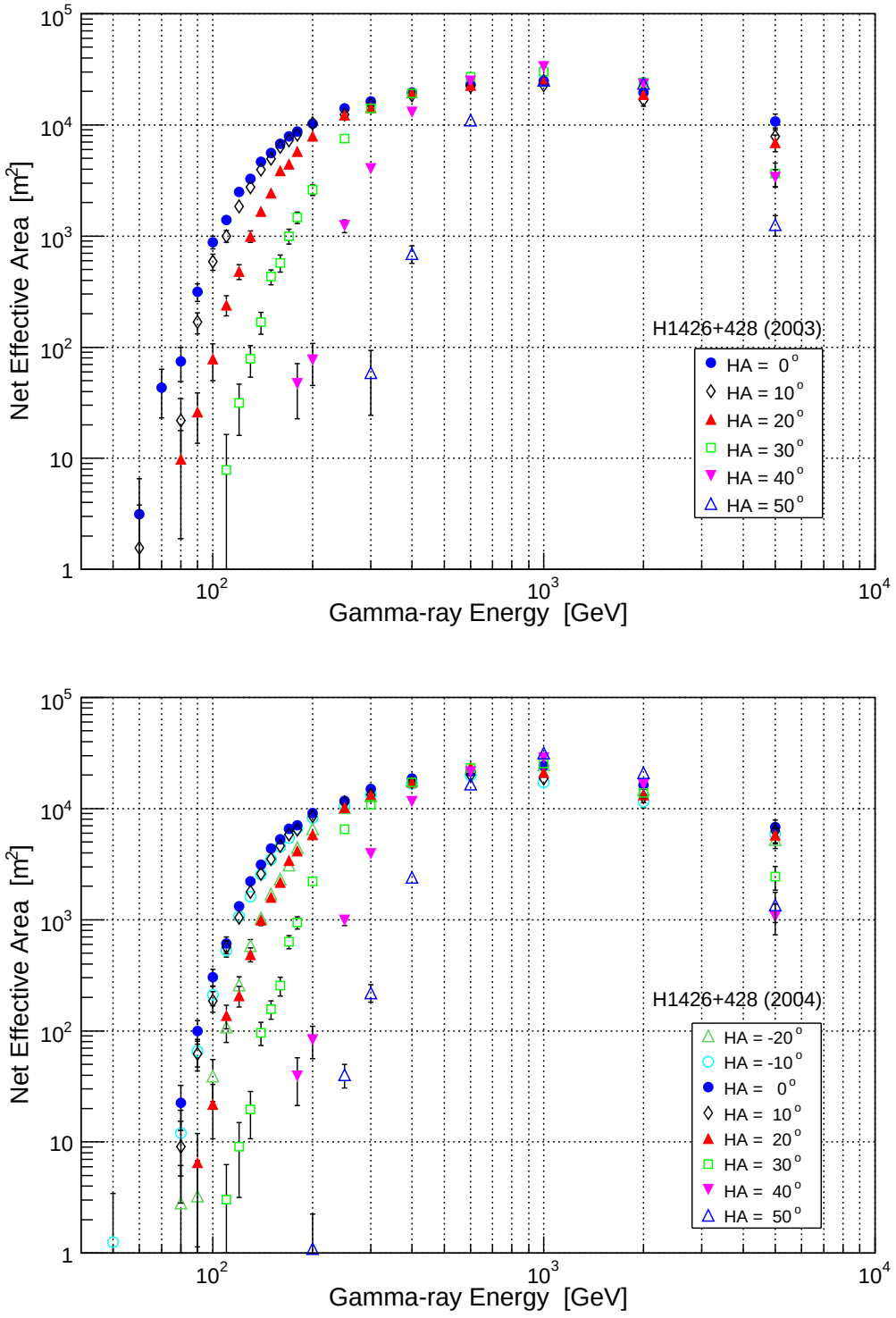

Figure 7.6: Net effective areas (*i.e.* after the application of event cuts) for the 2003 (*top*) and 2004 (*bottom*) observations of H 1426+428, for all 10° hour-angle (*HA*) bins for which there were observing time intervals passing the time cuts. Statistical errors are plotted.

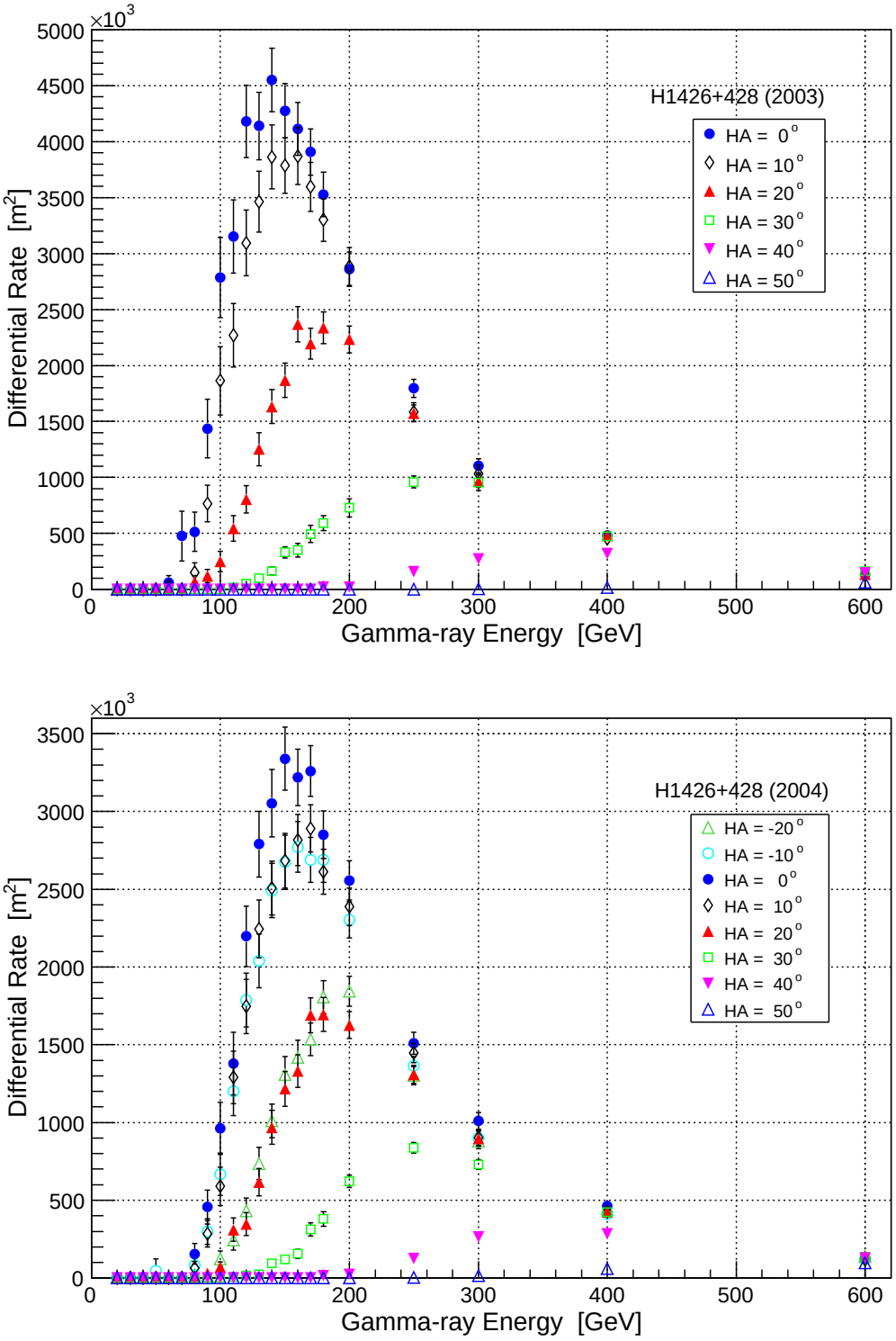

Figure 7.7: Detector response functions after the application of event cuts, for the 2003 (*top*) and 2004 (*bottom*) observations of H 1426+428, for all 10° hour-angle (*HA*) bins for which there was observing time available after the time cuts. The net effective areas have been folded with a power-law spectrum of index $\alpha = 3.5$, normalised to 1 TeV. The statistical errors are dominated by the uncertainties of the cut efficiency (see text).

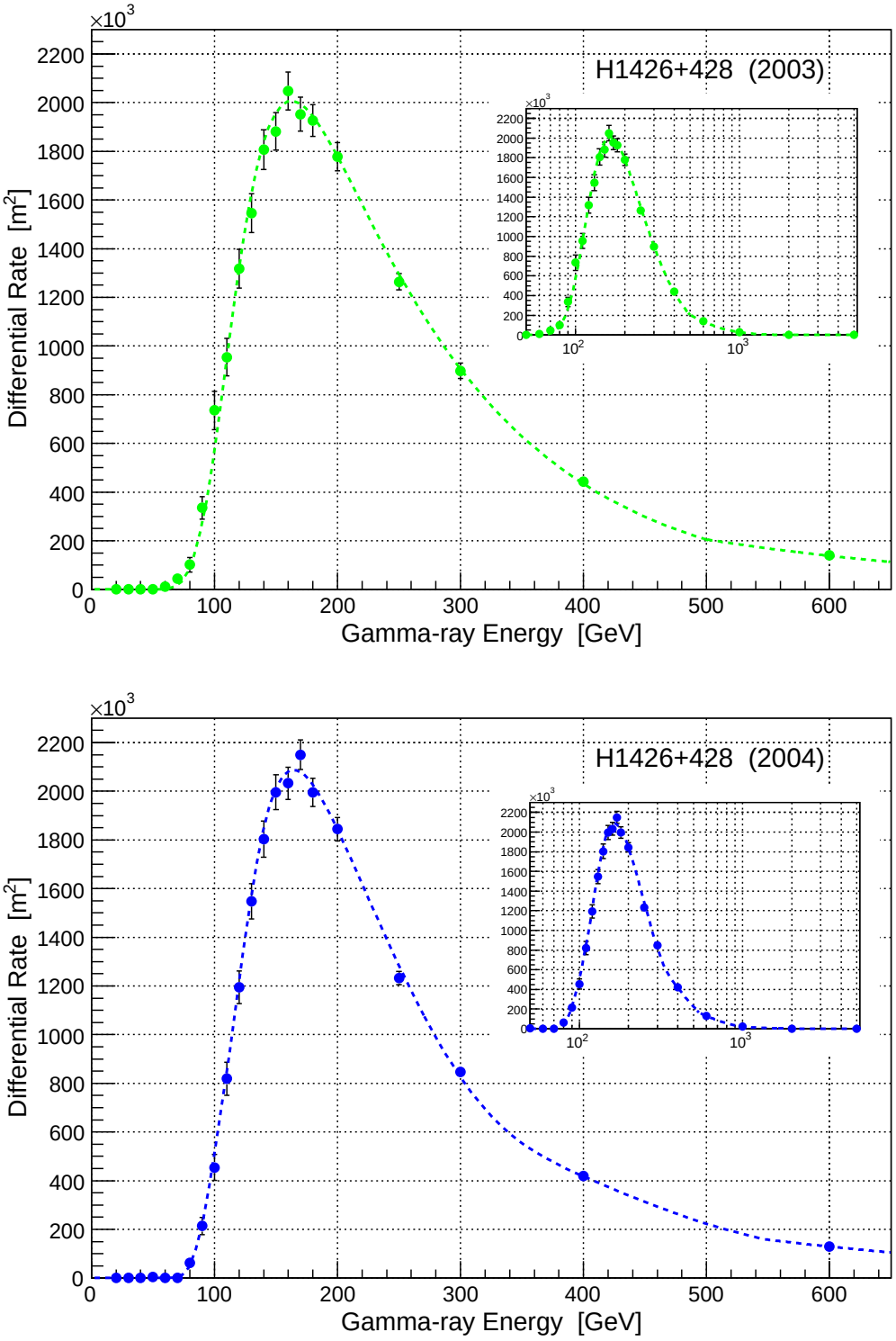

Figure 7.8: Final, exposure-averaged detector response functions for the 2003 (*top*) and 2004 (*bottom*) observations of H 1426+428. The net effective areas used for the *HA*-averaging were folded with a power-law spectrum of index $\alpha = 3.5$, normalised to 1 TeV. Statistical errors are shown. Empirical functions were fitted to the data. The same graphs, but with logarithmic energy scale, are shown in the inserts. The energy thresholds for the two observations are 163 GeV and 165 GeV, respectively.

The final detector response curves, exposure-averaged over the hour-angle distributions shown in Fig. 7.1, are plotted in Fig. 7.8. The data points have been fitted with empirical functions across the relevant energy range. It is interesting to note that even though the sensitivity was higher for the data taken with heliostats in monocanted configuration (*i.e.* for the observations of 2003), the exposure-averaged sensitivities of the two data sets are very similar. As we gather from comparing the two plots in Fig. 7.1, this is due to the zenith angles of the monocanted observations, which are on average much larger. The energy thresholds of the two data sets for the assumed spectral index ($\alpha = 3.5$) can be read off to be 163 GeV for 2003 and 165 GeV for 2004, with a statistical error of less than 5 GeV as obtained by varying the exact shapes of the curves around the peak. The detector responses are obtained by integrating the fitted curves, yielding the values $4.53 \cdot 10^5$ and $4.42 \cdot 10^5$ m^2 TeV for the 2003 and 2004 observations, respectively.

As an independent end-to-end check of STACEE's simulation predictions and data analysis procedure, we may use gamma-ray astronomy's "standard candle", the Crab Nebula. The latter is a constant and steady source of gamma rays across the whole energy band of interest to us. STACEE observed the Crab Nebula during the months prior to observations of H 1426+428 in both of the applicable observing seasons (2002-03 and 2003-04). The measured integral flux from the Crab Nebula above a threshold energy of 185 GeV was $(2.5 \pm 0.6) \cdot 10^{-6}\,\mathrm{m}^{-2}\mathrm{s}^{-1}$ for both data sets combined, in excellent agreement with spectral measurements at higher and lower energies (Fortin, 2005). We can therefore be confident that the gamma-ray fluxes calculated in the following section represent meaningful results.

7.4 Integral gamma-ray flux limits

Using the detector response integrals and energy thresholds provided by the detector simulation, we are able to proceed towards converting the observed excess rates to a measure of the gamma-ray flux. Evidently, since the final excess rates of both our data sets are consistent with zero, we must conclude that STACEE did not observe a gamma-ray signal from H 1426+428. Under these circumstances no actual flux can be calculated, but we are still able to quantify the non-observation of a gamma-ray excess by deriving an *upper limit* (UL) on the gamma-ray flux. Attached to the upper limit is a *confidence level* (CL), representing the probability that the actual flux is lower than the quoted UL. In gamma-ray astronomy, confidence intervals of 95% or 99% are commonly used.

7.4.1 Upper Limit calculation

Upper limits on the gamma-ray flux are derived via an upper limit on the excess event rate, determined by the measured signal and background rates. The upper limit on the rate is then coupled with the detector response predicted by the simulations and converted to a flux value by means of Eq. 7.5, where R_γ is replaced by the upper limit.

While the concept of setting an upper limit appears to be straightforward, the details of how upper limits should be calculated are still subject to significant controversy; this is particularly true when the measured value is close to the boundary of the physically allowed region, or even outside of it. Techniques differ mostly in the probability distribution used, and in the way it is normalised. For our purposes we here investigate two common methods.

Ordering principle of likelihood ratios

A recent solution to the problem of finding powerful confidence intervals that still provide the proper coverage was proposed by Feldman & Cousins (1998). In this publication, the authors attempt to address the problems by which existing techniques are frequently plagued. Conventional methods do not yield well defined confidence intervals or are overly conservative in the case of measurements very close to or beyond a physical boundary, *i.e.* in our case for zero or even negative excess rates. Bayesian approaches, on the other hand, do not have this problem; they are however affected by a certain degree of subjectivity, which, in the opinion of some authors, makes them unsuitable for a truly unbiased description of the data.

The method of Feldman & Cousins, a detailed explanation of which is beyond the scope of this work, provides a smooth transition between setting a confidence limit and stating a two-sided confidence interval, and it yields a well defined result in all cases. Applying the method is as simple as looking up the appropriate confidence intervals for the measured significances S_i of the two data sets in the tables for Gaussian likelihood distributions given by Feldman & Cousins (1998). For the significances of the two data sets, $S_{2003} = 1.56$ and $S_{2004} = 0.21$, the tables yield upper limits of S_{95} at the 95% confidence level of 3.52 and $2.17\,\sigma$, corresponding to rates $R_{\mathrm{UL},95} = S_{95} \cdot \sigma_{\mathrm{R}}$ of 1.44 and 0.59 min^{-1}, respectively. The σ_{R} are the errors on the rates as given in Table 7.2.

Unified Frequentist approach

Another standard method of calculating upper limits has been presented by the Particle Data Group (Barnett *et al.*, 1996). It involves renormalising the applicable likelihood

distribution to the physically allowed region. While upper limits calculated in this way may not be very powerful in the case of excess rates equal to or smaller than zero, the excess rates we calculated are positive. If we use the fact that our likelihood distribution is a Gaussian as before, the upper limit $R_{\mathrm{UL},\alpha}$ for a rate R with error σ_{R} is then defined by

$$\alpha \;=\; \frac{\int_0^{R_{\mathrm{UL},\alpha}} e^{-(r-R)^2/2\,\sigma_{\mathrm{R}}^2}\,\mathrm{d}r}{\int_0^{\infty} e^{-(r-R)^2/2\,\sigma_{\mathrm{R}}^2}\,\mathrm{d}r} \tag{7.6}$$

where α is the desired confidence limit, *i.e.* in our case $\alpha = 0.95$. Above equation can be solved for $R_{\mathrm{UL},\alpha}$ with the help of the Error function *erf()*, which satisfies the identity

$$\mathrm{erf}\Big(\frac{\delta}{\sqrt{2}\sigma}\Big) \;=\; \int_{\mu-\delta}^{\mu+\delta} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}\,\mathrm{d}x \tag{7.7}$$

yielding the result

$$R_{\mathrm{UL},\alpha} \;=\; R \;+\; \sqrt{2}\,\sigma_{\mathrm{R}}\,\mathrm{erf}^{-1}\Big[\alpha \;+\; (\alpha-1)\,\mathrm{erf}\Big(\frac{R}{\sqrt{2}\,\sigma_{\mathrm{R}}}\Big)\Big] \tag{7.8}$$

which can readily be evaluated (erf^{-1} is the inverse Error function). Substituting the appropriate values for R and σ_{R}, Eq. 7.8 gives the upper limits $R_{\mathrm{UL},95}$ of 1.34 and 0.57 min^{-1} for the 2003 and 2004 data sets, respectively.

It does not come entirely unexpected that the limits obtained with the conventional approach are slightly more restrictive than those predicted by the more advanced method, as the excess rates are non-negative for both data sets. We therefore choose to use the Unified Frequentist approach. Converting these rate limits to upper limits on the gamma-ray flux via Eq. 7.5, we obtain 95% CL limits on the flux normalisations $N_{0,\mathrm{UL}}$ of $4.92 \cdot 10^{-8}$ and $2.15 \cdot 10^{-8}$ $\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{TeV}^{-1}$ for the 2003 and 2004 data sets, respectively.

Integral and differential flux limits

The upper limits on the flux normalisation translate into upper limits on the integral fluxes above the threshold energies E_{T}. For the assumed power-law spectrum with index $\alpha = 3.5$, the integral fluxes are

$$\begin{aligned}\Phi_{\mathrm{int}}(E > E_{\mathrm{T}}) \;\equiv\; \int_{E_{\mathrm{T}}}^{\infty} \frac{\mathrm{d}N}{\mathrm{d}E}\,\mathrm{d}E \;&=\; N_{0,\mathrm{UL}} \int_{E_{\mathrm{T}}}^{\infty} \left(\frac{E}{E_0}\right)^{-\alpha} \mathrm{d}E \\ &=\; N_{0,\mathrm{UL}}\,\frac{1}{\alpha-1}\,E_0^{\alpha}\,E_{\mathrm{T}}^{1-\alpha}\end{aligned} \tag{7.9}$$

where $E_0 = 1$ TeV as before. Substituting the derived values for $N_{0,\mathrm{UL}}$, one obtains the upper limits (95% CL) on the integral fluxes

$$\begin{aligned} 2003 \text{ observ. :} \quad & \Phi_{\mathrm{int}}(E > 163\,\mathrm{GeV}) < 1.83 \cdot 10^{-6}\ \mathrm{m^{-2}\,s^{-1}} \\ 2004 \text{ observ. :} \quad & \Phi_{\mathrm{int}}(E > 165\,\mathrm{GeV}) < 0.78 \cdot 10^{-6}\ \mathrm{m^{-2}\,s^{-1}} \end{aligned}$$

In order to compare these upper limits with existing measurements, one must first convert them to differential fluxes. As the obtained upper limits only apply to the *integral* flux, this is only possible if the spectral shape of the SED across the energy range in which STACEE has significant sensitivity is indeed a power law. Since for the spectral index α, STACEE attains its maximum sensitivity at the spectral energy threshold (see Fig. 7.8), the latter is the energy that most accurately describes the measurement. In this case, the corresponding differential flux can therefore be calculated by evaluating Eq. 7.3 at the threshold energy E_{T}. As upper limits on the flux at these energies we then obtain for the two observing seasons:

$$\begin{aligned} 2003 \text{ observ. :} \quad & \left(\frac{\mathrm{d}N}{\mathrm{d}E}\right)_{\mathrm{int}} (E > 163\,\mathrm{GeV}) < 2.81 \cdot 10^{-5}\ \mathrm{m^{-2}\,s^{-1}\,TeV^{-1}} \\ 2004 \text{ observ. :} \quad & \left(\frac{\mathrm{d}N}{\mathrm{d}E}\right)_{\mathrm{int}} (E > 165\,\mathrm{GeV}) < 1.18 \cdot 10^{-5}\ \mathrm{m^{-2}\,s^{-1}\,TeV^{-1}} \end{aligned}$$

When plotting steeply falling flux spectra, the flux values are usually multiplied by the square of the energy in order to enhance the visibility of spectral variations. Converting the upper flux limits to these new units (and also using centimetres as units of length, as is common in gamma-ray astrophysical publications), one obtains $7.46 \cdot 10^{-11}$ and $3.21 \cdot 10^{-11}$ TeV cm^{-2} s^{-1} for the 2003 and 2004 observations, respectively. How these upper limits relate to other gamma-ray flux measurements in the VHE-energy range will be examined in Sec. 7.5.

7.4.2 Sources of systematic error

The calculation of the upper flux limits is based on statistical errors resulting from the rate measurements themselves, and hence does not account for additional, systematic errors in the measurement. While the technique used to calculate the rates is certainly affected by intrinsic uncertainties – *e.g.* errors introduced by the software padding scheme – all the procedural errors are small compared to the uncertainties adherent to the detector calibration, affecting the various stages of the detector simulation. Uncertainties

in the simulations affect both energy scale and flux values, and are mostly a result of the relatively large errors on the total throughout of the detector.

The throughput is essentially the transfer function between the primary gamma-ray energy and the number of photoelectrons released from the PMT photocathodes. It depends on quantities such as Cherenkov light yield, atmospheric attenuation, throughput of the STACEE optics, quantum efficiencies, PMT gains and discriminator thresholds. Simulations indicate that other possible sources of error such as differences between the degree of fluctuation of real and simulated Cherenkov shower timing profiles, channel-to-channel fluctuations of the throughput *etc.* do not measurably influence the overall response to gamma-ray showers, and that the systematic errors are dominated by the *mean* behaviour of the detector. The output of the detector simulations is represented by the effective area curves, which are subject to systematic errors in both energy scale and absolute value. Of the latter two, only the error in the energy scale is of relevance, because the total achievable effective area itself depends mainly on the geometry of the heliostat field, which is well known.

In order to quantify the effects of the uncertainty in the energy scale caused by the errors in the overall throughput, one needs to consider how the trigger decision is affected by the various uncertain quantities as a function of energy. Fortunately, the calorimetric nature of Cherenkov light facilitates error analysis significantly. In Sec. 3.1.2 we saw that the Cherenkov light density on the ground scales approximately linearly with the primary gamma-ray energy. All other physical properties of the resulting Cherenkov light showers are, for the purpose of this argument, identical: the diameter of the illuminated area on the ground does not depend on the shower energy, and the slight dependence of the wavefront shape on gamma-ray energy is too small to affect the trigger mechanism in any way. Any change in throughput will thus affect the energy scale of the experiment in a quasi linear fashion.

Mathematically, this can be formulated as follows. If $A_{sim}(E)$ is the effective area function of the detector as predicted by the simulations, and $A(E)$ the (unknown) actual effective area, the two are related via

$$A(E) \;=\; A_{sim}(\eta E) \tag{7.10}$$

Here $\eta \equiv E/E_{sim}$ is the ratio of the true to the apparent energy of a gamma-ray shower, which, from what was explained above, is equivalent to the ratio of the real to the predicted throughput. Any deviation of η from unity thus corresponds to an error in the

overall calibration of the detector. The combined fractional uncertainties in the energy scale, $\Delta\eta$, do then satisfy the condition $\eta = 1 \pm \Delta\eta$.

It is intuitive that an error in the energy scale of the effective area should result in a corresponding error in the energy threshold $E_{\rm T}$ of the detector. We recall that the energy threshold is defined as the energy at which the product of effective area and differential flux is maximised, and that for a power-law spectrum, this product is

$$A(\eta E)\, E^{-\alpha} \tag{7.11}$$

where the effective area function may be shifted to lower or higher energies due to a misestimation of the true shower energy. Substituting $E_{sim} = \eta E$, the above expression becomes

$$A(E_{sim})\, (E_{sim}/\eta)^{-\alpha} \;=\; A(E_{sim})\, E_{sim}^{-\alpha}\, \eta^{\alpha} \tag{7.12}$$

which is the same functional form as Eq. 7.11, and hence maximised for the *same* value of the argument of $A(E)$. However, since $E_{sim} = \eta E$, the respective threshold energies are related in the same manner, and it follows that

$$E_{\rm T} \;=\; \frac{E_{{\rm T},sim}}{\eta} \;\approx\; E_{{\rm T},sim}\,(1 \mp \Delta\eta) \tag{7.13}$$

where the right-hand approximation is the first-order Taylor expansion, valid under the assumption that $\Delta\eta \ll 1$. What Eq. 7.13 tells us is that the linearity condition is indeed fulfilled for power-law spectra, and the fractional uncertainty of the energy threshold is identical to the fractional uncertainty in the energy scale of the effective area. In order to investigate the effect of errors in the energy scale on the integrated flux (limit), one must replace $A(E)$ with $A(\eta E)$ and E_T with E_T/η in Equations 7.5 and 7.9. By doing so, one can easily show that:

$$\Phi_{\rm int}(E > E_{{\rm T},sim}) \;=\; \Phi_{\rm int}(E > E_{\rm T} = E_{{\rm T},sim}/\eta) \tag{7.14}$$

The two factors of η have conveniently cancelled out, which means that for a power-law spectrum, the throughput parameter η changes the energy threshold, but it has no effect on the calculated integral flux value.

Rather than investigating the effects of individual uncertainties separately and propagating the errors through the complete simulation, one may thus simply add the individual errors $\Delta\eta$ in quadrature, and so obtain the total systematic uncertainty in the energy scale.

Breakdown of systematic uncertainties

The various sources of systematic error in the overall energy calibration of STACEE are estimated as follows:

1. **Cherenkov light production:** By comparing the results of different air shower simulation codes, one may obtain an estimate of the uncertainty in the predicted intensity of Cherenkov light produced by gamma-ray air showers of a given energy. The physics of electromagnetic air showers are well understood, in particular at the energies relevant for STACEE, hence this error is small. Comparisons between the CORSIKA and MOCCA codes suggest that it is at most **5%** (see Oser (2000) and Sec. 5.5.1).

2. **Atmospheric attenuation:** In Sec. 5.2.3 we compared the parameterization of the atmosphere we used to alternative atmospheric models, and found that the cosmic-ray trigger rates differed by maximally 10%. The dependence of the trigger rates on the Cherenkov light density on the ground is more complicated, but one may safely assume that they vary with the densities by some power larger than one. Hence, using **10%** as the error due to atmospheric transparency is a rather conservative choice.

3. **Optical alignment of heliostats and secondary mirrors:** Owing to the various optical calibration methods described in Sec. 4.2, the errors due to heliostat defocusing and misalignment have been minimised considerably; an uncertainty of about 2 pointing bits in the heliostat biasing remains. Combined with the degree of heliostat facet misalignment, which was measured only infrequently, we estimate a total error of **5%** in the optical alignment of the heliostats. This figure seems to be supported by sunspot measurements (see Sec. 5.5.2). The degree of misalignment of the secondary mirrors is too small to affect the light collection in a measurable way.

4. **Heliostat reflectivity:** The heliostat reflectivity was measured as a function of wavelength by means of a sample of heliostat glass. While this measurement was probably quite accurate, it dates to the days when the experiment was constructed (Oser, 2000). Between 1998 and 2003/2004, weathering of the heliostat facets can be expected to have degraded the reflectivity slightly. We assign an error of **7%** to the heliostat reflectivity.

5. **Secondary reflectivity:** Monthly reflectometer measurements were performed to monitor the reflectivity of the secondary mirrors. Individual measurements across

the facet surface differed by about 5%, hence we conservatively assume an error of **5%** for the secondary reflectivities.

6. **PMT can misalignment:** From intensity distributions of CCD camera spots and the way in which the positions of individual PMT cans are aligned in the simulation (they are placed at the theoretically best aligned positions, which due to the enhancement effect however do not necessarily correspond to the positions of optimum light collection), one may assume an uncertainty of **5%** in the light collection of the DTIRCs due to can position.

7. **Quantum efficiency:** According to the manufacturer, the quantum efficiency may vary up to 13% between individual PMTs (Photonis PMT catalogue, 2004; www-Photonis, 2005). The corresponding *mean* error of a sample of 64 phototubes is about 2%. Since we do not measure the quantum efficiency alone, but the combined effects of DTIRC, cookie and PMT, the uncertainty will be larger; laboratory tests of the assembled cans (see Sec. 4.1.3) revealed a scatter of about 10% between individual cans, compatible with the above error on the quantum efficiency. As we are only interested in the mean uncertainties, which are much smaller, we assume the error on the quantum efficiency and DTIRC transmission to be **5%**.

8. **PMT gains:** As a result of the limited accuracy of the gain measurement techniques and the tendency of the gains to drift over time, one may assume that they are affected by a significant error. Yet, while the channel-to-channel uncertainty has been observed to be about 15%, the mean error is most likely much smaller. Considering that the PMT rates can be reproduced very accurately by the electronics simulation (Sec. 5.5.4), the average gains cannot be off by more than a few percent. The assumption that the mean PMT gain has an error of **5%** is therefore conservative.

Since all of the above errors are uncorrelated, they can be added in quadrature; the result is $\Delta\eta = 17\%$. The spectral threshold energies to be quoted are therefore:

$$\begin{aligned} &\text{2003 observations:} \quad E_{\mathrm{T}} = (163 \pm 28)\ \text{GeV} \\ &\text{2004 observations:} \quad E_{\mathrm{T}} = (165 \pm 28)\ \text{GeV} \end{aligned}$$

In addition to the (dominant) errors on the energy scale, there are sources of error that do affect the values of the calculated (integral and differential) flux limits. The above considerations apply to the case of a power-law spectrum with known spectral

index α. Departures of the actual shape of the SED from the assumed power law, and in particular uncertainties on α – in part due to the uncertainties of the spectral measurements performed by the imaging ACTs which were taken as a basis for the choice of spectral index to be used, in part because the actual value of α may be somewhat different in the (lower) energy range applicable to STACEE – will affect energy threshold and the response function integral in a way that does not lead to a mutual cancellation of both effects. In order to investigate the effects of using different spectral shapes, response functions were obtained by folding the detector's effective areas with power laws of index $\alpha \pm \Delta\alpha$, where $\Delta\alpha = 0.25$ was used. The choice of $\Delta\alpha$ was motivated by the errors in the spectral measurements of the IACTs, and by an estimate of the possible curvature of the spectrum towards lower energies as observed in the SEDs of other blazars. The resulting upper flux limits differ from our nominal results by $\sim 10\%$, and the energy threshold varies by less than 4%. The additional error on the energy threshold is small compared to the uncertainties affecting the optical throughput, and may therefore be neglected. Upper flux limits for the cases of $\alpha = 3.25$ and $\alpha = 3.75$ must be quoted at the modified energy thresholds, which are shifted by ~ 5 GeV with respect to the quoted thresholds. The resulting points lie almost right on a power law of spectral index $\alpha = 3.5$ through the nominal upper limits; combined with the small magnitude of this error, we can thus neglect the effect of the uncertainty in the spectral index.

7.5 Astrophysical Implications

The upper flux limits derived in Sec. 7.4.1 are plotted in Fig. 7.9, along with results of gamma-ray observations of H 1426+428 at higher energies. The uncertainties in the energy scale calculated in the previous section are indicated in the form of horizontal error bars on the two STACEE upper limit points. In order to interpret Fig. 7.9, one must bear in mind that most of the other flux measurements are not contemporaneous, although the observations of Whipple and CAT do overlap with the HEGRA measurements of 1999/2000. During this epoch of 1998 - 2001, the blazar appears to have remained in a relatively high state of emission. The solid line shown in the plot is a power-law fit to the energy spectrum measured by Whipple in 2001 (Petry *et al.*, 2002). As noted by Aharonian *et al.* (2003), subsequent observations of the source by HEGRA revealed a significantly lower flux level in the spring of 2002, indicated by the cyan circles in Fig. 7.9. The upper limit we derived for the STACEE observations of H 1426+428 in 2003 is not very powerful, yet it constrains the source flux for that epoch to be no higher

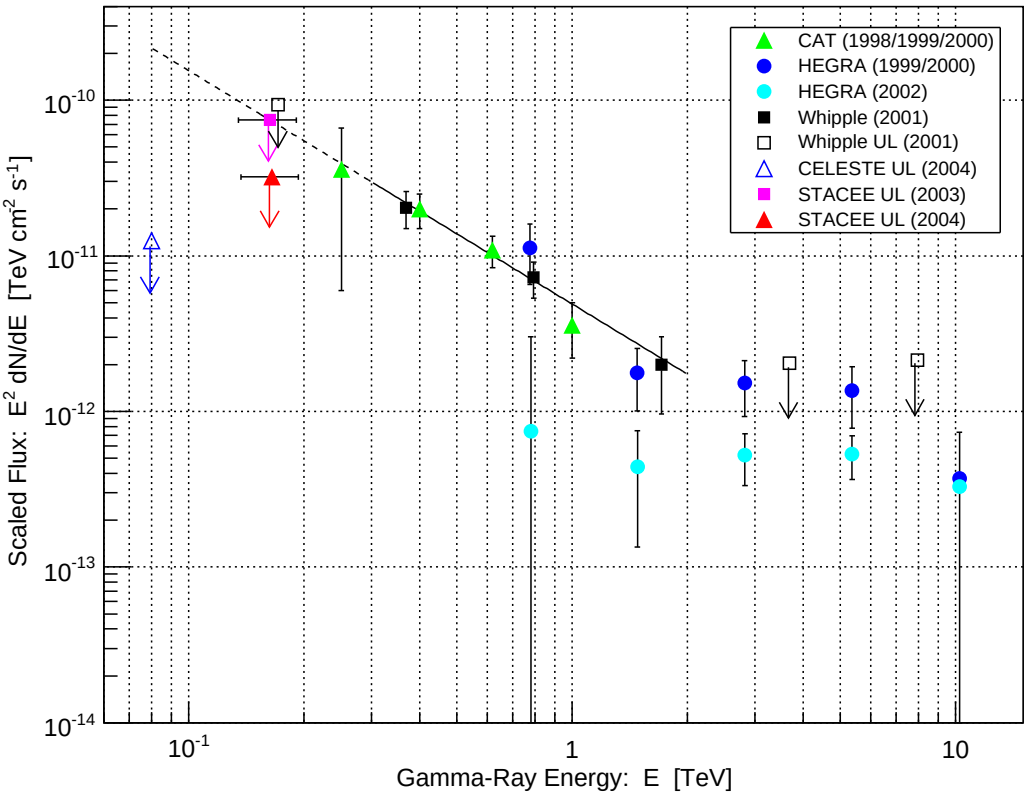

Figure 7.9: Measurements of the differential gamma-ray flux from H 1426+428 at VHE energies, over several observing seasons. Upper flux limits (UL) are denoted by downward pointing arrows. The STACEE upper limits at the 95% CL obtained in this work (2003: *pink square*; 2004: *red triangle*) and the CELESTE upper limit of 2004 (*empty blue triangle*) are limits on the integral flux converted to differential flux limits; for the STACEE measurements, systematic errors in the energy scale are indicated. The flux state of the blazar appears to have remained approximately constant between 1998 and 2001, as the measurements by CAT (*green triangles*), Whipple (*black squares*) and the overlapping observations by HEGRA (*blue circles*) suggest. The solid black line is a power-law fit to the Whipple points (Petry *et al.*, 2002), intended to describe the SED at energies between $\sim$300 and 2000 GeV during the 1999-2001 observing seasons; the dashed line denotes the continuation of this power law towards lower energies. In 2002 the blazar appeared to be in a state of lower activity, which is reflected in the lower fluxes measured by HEGRA in that epoch (*cyan circles*). A possible further flux decrease over the subsequent years may explain the non-detection of the source by the two wavefront-sampler experiments. References for all results are given in Sec. 2.6.2.

than the flux level observed by the imaging ACTs in the period of 1998 - 2001, under the assumption that the flux in energy range around 160 GeV can be described by the same power law. The reason for the relative weakness of this limit is the fact that STACEE observed an on-source excess in this data set, albeit not a statistically significant one. Whether the observed excess is due to a statistical fluctuation or due to actual gamma rays is of course a mere matter of speculation.

The upper limit we derived for STACEE's observations during the 2004 observing season is more powerful. A continuation of the power law describing the TeV-observations during the 1998 - 2001 period to lower energies (dashed line in Fig. 7.9) can clearly be excluded. This may have either of two reasons: a significant decrease of the overall flux – if one assumes a uniform flux decrease (as opposed to a flux decrease coupled with a significant spectral softening), this UL appears to agree with the fluxes measured by HEGRA in 2002 – or the fact that the (post-absorption) IC-peak is located around 250 - 300 GeV. The latter would have to be rather narrow, a result of EBL-absorption becoming significant beyond about 200 GeV. Even for a relatively far gamma-ray source like H 1426+428, attenuation by the extragalactic background light is almost negligible at the threshold energies we derived for STACEE (see Fig. 2.12), however as the gamma-ray energy approaches 1 TeV, the likelihood of pair-production increases sharply. If we were to assume a constant emission, the inferred narrow IC-peak around $\sim$300 GeV would hence provide a constraint on combinations of EBL-models and intrinsic emission spectra. Unfortunately, since H 1426+428 is known to undergo large-amplitude, long-term flux variations, our upper limits fail to constrain the spectrum of the extragalactic near-infrared background light (see also Sec. 2.5). For the same reasons, we also cannot make a statement in regard to the presence or absence of secondary (cascade) gamma-ray emission from the direction of the source.

The limit imposed by the results of CELESTE's observations in 2004 appears to be more restrictive than STACEE's ULs, as it is more than an order of magnitude below the extrapolation of the ($E > 300$ GeV) fluxes measured in the period of 1998 - 2001.[8] It does not, however, provide a constraint on the EBL density, as no existing blazar emission model could account for the inferred steep rise of the SED in the GeV-range. This fact again argues for our hypothesis that H 1426+428 was in a state of rather low activity during the spring of 2004.

The broad-band SED of H 1426+428 was shown in Fig. 2.19; EGRET only published an upper limit, hence the SED is only weakly constrained in the MeV energy range. As already argued in Sec. 2.6.3, however, X-ray observations suggest that at least during

[8] No details are known about the data supporting this upper limit, nor about the confidence level.

high-flux states, the (intrinsic) Inverse-Compton peak of the blazar is located at energies well above STACEE's energy range.

7.5.1 Contemporaneous X-ray observations

While there are no targeted, sensitive X-ray observations of H 1426+428 available that temporally overlap with the two STACEE observations, the ASM detector onboard RXTE (Levine *et al.*, 1996), already discussed in Sec. 2.6.4, routinely monitored the sky region around H 1426+428 during both periods. ASM is able to detect X-rays between 1.5 and 12 keV and has an angular resolution of 3'×15'; unfortunately, the low exposure times characteristic of an all-sky monitoring device result in a relatively low sensitivity.

All ASM data can be downloaded in tabulated form from the RXTE web site (www-RXTE, 2007). Due to the low ASM count rates for the solid-angle bin containing H 1426+428, we chose to download the daily averages for this source and, for both the 2003 and 2004 observations, calculated the average rate for the full interval during which observations with STACEE were made, *i.e.* using all ASM-rates from the first to the last UT-date of the respective STACEE observation period. Both count rates were then normalised to the contemporaneous ASM-rates from the direction of the Crab Nebula[9], which fluctuate around 75 Hz. We then arrive at the following rates R_{ASM}:

$$\begin{aligned} &\text{2003 observations:} \quad R_{\mathrm{ASM}} = (1.80 \pm 0.74)\ \text{mCrab} \\ &\text{2004 observations:} \quad R_{\mathrm{ASM}} = (0.22 \pm 0.88)\ \text{mCrab} \end{aligned}$$

Even though the errors are large and both numbers are still (barely) compatible with each other, these rates do suggest that the X-ray flux from the direction of H 1426+428 was higher during the spring of 2003 than during the following spring. If one expects a *positive* correlation between X-ray and GeV gamma-ray fluxes, as suggested by most spectral models, this would support our hypothesis that the excess observed by STACEE in the 2003 period is indeed due to gamma rays.

It should be noted that the same trend is seen when including only those ASM observations in the average that coincided with the UT-days, or even exact hours, of STACEE observations. In these cases however the statistical errors are so large that no quantitative conclusions can be drawn.

[9]The Crab Nebula is a steady, non-variable source of radiation also in X-rays.

7.6 Conclusion

The STACEE gamma-ray detector did not produce any ground-breaking results during its life time, such as a detection of gamma-ray emission from previously unknown sources. It did however provide valuable results that complement those of established gamma-ray telescopes, in an energy range that is only gradually being explored by the new generation of imaging atmospheric Cherenkov telescopes. Blazars are highly variable radiation sources by nature, and in order to get to the bottom of the phenomena driving these powerful, mysterious objects, sufficient temporal coverage is almost as important as are observations over a large spectral range. Every observation that gives rise to a physically meaningful result therefore adds to our knowledge base about the particular observed source, and this class of astrophysical objects in general.

We have here derived upper limits on the integral flux of gamma rays from the direction of H 1426+428 for two subsequent observing seasons. Given STACEE's sensitivity to gamma rays, which we have shown to be sufficiently well known by means of measurements of the gamma-ray flux from the Crab Nebula, these upper limits are surprisingly competitive. We have exploited the absence of grave background normalisation problems in the case of this particular observation target. Had STACEE been available for observing the blazar during its time of high activity in 2000/2001, it would likely have accumulated a statistically significant signal due to gamma rays from this source.

The derived upper limits do not constrain the level of extragalactic infrared background light, as H 1426+428 is known to be a highly variable gamma-ray source, as are all BL Lac objects. We can however conclude that the blazar was in a state of relatively low activity during the spring of 2004, a claim that is also supported by the upper limit published by the CELESTE collaboration for observations during the same period.

Also due to the lack of contemporaneous reference data, the presented upper limits do not shed any light on the existence of the proposed secondary gamma-ray emission in the GeV range, caused by pair-production of TeV gamma rays on the CMBR. Due to the transient nature of blazars, multi-wavelength observations are needed to address the interesting topics of intergalactic background light and cascade emission.

In conclusion it shall be pointed out that despite the increasingly large number of discovered extragalactic VHE gamma-ray sources at distances larger than that of H 1426+428, the latter is still a worthwhile target for multi-wavelength campaigns – last but not least because the X-ray radiation it emits flags it as one of only a few known astrophysical sources that are capable of accelerating particles to energies higher than any other objects in the hitherto observed universe.

APPENDIX A

SIMULATION PARAMETERS

A.1 CORSIKA input parameters

The CORSIKA air-shower simulation input card, one of which is needed for each shower to be generated, consists of a list of *keywords*, an associated (array of) value(s) and a comment field, one of each on every line. The keyword "`EXIT`" on the last line terminates the input. A typical input card for the STACEE version of CORSIKA – here the version using FLUKA and QGSJET as hadronic models – is given in the following Table A.1.

```
RUNNR   13579                       number of run
EVTNR   1                           number of first shower event
NSHOW   1                           number of showers to generate
PRMPAR  14                          primary particle code
ESLOPE  -2.7                        slope of primary energy spectrum
ERANGE  107.196308  107.196308      energy range of primary particle
THETAP  1.866989 1.866989           range of zenith angle (degree)
PHIP    128.767792 128.767792       range of azimuth angle (degree)
SEED    6386   0   0                seed for hadronic part
SEED    7940   0   0                seed for EGS4 part
SEED    4523   0   0                seed for Cherenkov part
OBSLEV  1705.0E2                    observation level (cm)
ELMFLG  F  T                        em. interaction flags (NKG, EGS)
STEPFC  1.0                         mult. scattering step length fact.
RADNKG  1000.E2                     outer radius for NKG lat.dens.determ.
FIXHEI  0  0                        first interaction height (cm) & target
FIXCHI  0.                          starting altitude (g/cm^2)
MAGNET  23.55 44.98                 magnetic field ALBUQUERQUE 2004 (/uT)
HADFLG  0  0  0  0  0  2            flags for hadr.interact.&fragmentation
QGSJET  T  0                        use QGS for high energy hadrons
QGSSIG  T                           use QGS hadronic cross-sections
ECUTS   0.1  0.02  0.02  0.02       energy cuts for particles (GeV)
MUADDI  F                           additional muon information
MUMULT  T                           muon multiple scattering by Moliere
LONGI   F  20.  F   F               longit.distr., step size, fit & out
MAXPRT  1                           max. number of printed events
ECTMAP  1.E0                        cut on gamma factor for printout
THIN    1.E-9   1.E8   0.E0         efrcthn, wmax, rmax (in cm)
ATMOD   1                           atmospheric model selection
ARRANG  -79.79                      rotation mag.north to array grid (deg)
CERARY  1 1 1. 1. 262000. 262000.   def. of Cherenkov array grid (cm)
CERSIZ  1.                          bunch size Cherenkov photons
CERFIL  F                           Cherenkov output to extra file
CWAVLG  200.  700.                  Cherenkov wavelength band (nm)
CSCAT   1  0. 0.                    scatter Cherenkov events (cm)
CERQEF  T  T  F                     Cherenkov efficiency and absorption
HILOW   100.                        transition energy between models (GeV)
DIRECT  /dev/null                   output directory
DEBUG   F  9  F  100000             debug flag, log. unit, delayed debug
EXIT
```

Table A.1: Typical input file to steer the CORSIKA air-shower simulation (STACEE version). For an explanation of the parameters, see the comments on the right.

A.2 Optical simulation parameters

The channel-by-channel section of a typical input file for the optical simulation `Sandfield` is listed in the following Tables A.2 and A.3, containing heliostat positions and alignment parameters. The canting parameter values shown are for the *paracanting* scheme. Heliostat coordinates are expressed in metres relative to the base of the Solar Tower, where the x-axis points East and the y-axis North. The detector centre (average of all heliostat positions) is at $(x, y, z) = (0.005, 117.092, 4.34)$ m. For all simulations, heliostat azimuth and elevation bias errors were assumed to be zero for all channels.

Chan.	Helio-ID	x [m]	y [m]	z [m]	Sec.	σ_f	Cant.	ξ [°]	PMT-No.
00	07E07	63.39	82.28	5.36	0	0.2	0.00	19.0	19325
01	08E06	53.62	94.48	5.32	0	0.2	0.08	19.0	19342
02	08E09	82.87	94.48	5.92	0	0.2	0.08	24.0	19584
03	10E08	73.13	121.59	6.07	0	0.2	0.00	24.0	19522
04	08E04	34.11	94.50	4.96	0	0.2	0.08	19.0	19556
05	09E03	24.39	107.29	4.92	0	0.2	0.08	19.0	19336
06	09E05	43.88	107.29	5.29	0	0.2	0.08	19.0	19463
07	10E04	34.12	121.59	5.29	0	0.2	0.08	24.0	19332
08	11E07	63.38	136.85	5.94	0	0.2	0.08	24.0	19437
09	12E05	43.87	154.53	5.54	0	0.2	0.08	24.0	19873
10	12E09	82.87	154.52	6.46	0	0.2	0.08	28.1	19591
11	13E07	63.38	173.71	6.00	0	0.2	0.00	28.1	19596
12	10E10	92.60	121.63	6.45	0	0.2	0.08	24.0	19362
13	11E06	53.63	136.84	5.74	0	0.2	0.00	24.0	19335
14	12E12	112.14	154.51	7.16	0	0.2	0.08	28.1	19249
15	13E12	112.12	173.72	7.17	0	0.2	0.08	28.1	19582
16	07W07	-63.37	82.30	2.92	1	0.2	0.00	19.0	19581
17	08W06	-53.60	94.48	3.17	1	0.2	0.08	19.0	19351
18	08W09	-82.85	94.49	2.47	1	0.2	0.08	24.0	19453
19	10W08	-73.14	121.62	2.79	1	0.2	0.00	24.0	19557
20	08W04	-34.12	94.49	3.62	1	0.2	0.08	19.0	19363
21	09W03	-24.37	107.29	3.90	1	0.2	0.08	19.0	19358
22	09W05	-43.86	107.29	3.40	1	0.2	0.08	19.0	19560
23	10W04	-34.11	121.61	3.71	1	0.2	0.08	24.0	19597
24	11W07	-63.38	136.86	3.01	1	0.2	0.08	24.0	19588
25	12W05	-43.88	154.52	3.48	1	0.2	0.08	24.0	19329
26	12W09	-82.89	154.52	2.55	1	0.2	0.08	28.1	19587
27	13W07	-63.38	173.75	3.07	1	0.2	0.00	28.1	19578
28	10W10	-92.62	121.63	2.33	1	0.2	0.08	24.0	19589
29	11W06	-53.62	136.85	3.22	1	0.2	0.00	24.0	19451
30	12W12	-112.13	154.54	1.85	1	0.2	0.08	28.1	19455
31	13W12	-112.12	173.72	1.92	1	0.2	0.08	28.1	19561

Table A.2: Heliostat positions and channel-specific `Sandfield` parameters (*part I*).

Chan.	Helio-ID	x [m]	y [m]	z [m]	Sec.	σ_f	Cant.	ξ [°]	PMT-No.
32	08E02	14.63	94.49	4.53	2	0.2	0.08	19.0	19460
33	08W02	-14.63	94.49	4.02	2	0.2	0.08	19.0	19345
34	09E01	4.89	107.30	4.48	2	0.2	0.00	19.0	19486
35	10E02	14.62	121.60	4.83	2	0.2	0.08	24.0	19355
36	10W01	-4.86	121.62	4.36	2	0.2	0.08	19.0	19348
37	10W03	-24.35	121.61	3.92	2	0.2	0.00	19.0	19356
38	11E04	34.13	136.84	5.27	2	0.2	0.08	24.0	19311
39	11E01	4.88	136.84	4.59	2	0.2	0.08	24.0	19583
40	11W02	-14.63	136.85	4.18	2	0.2	0.08	24.0	19324
41	12E03	24.39	154.52	5.09	2	0.2	0.08	24.0	19343
42	12W01	-4.89	154.52	4.44	2	0.2	0.00	24.0	19549
43	13E02	14.62	173.72	4.89	2	0.2	0.08	28.1	19590
44	13W03	-24.36	173.73	4.02	2	0.2	0.08	28.1	19866
45	14E04	34.10	194.74	5.41	2	0.2	0.08	28.1	19874
46	14W02	-14.65	194.74	4.31	2	0.2	0.00	28.1	19867
47	14W05	-43.88	194.75	3.59	2	0.2	0.08	28.1	19594
48	05E04	34.12	57.94	4.44	3	0.1	0.08	24.0	19340
49	05E03	24.37	57.93	4.28	3	0.1	0.08	24.0	19333
50	05E02	14.65	57.94	4.15	3	0.1	0.00	24.0	19310
51	05E01	4.89	57.93	4.02	3	0.1	0.08	24.0	19593
52	07E05	43.89	82.29	4.97	3	0.1	0.08	28.1	19548
53	07E04	34.13	82.30	4.78	3	0.1	0.00	28.1	19862
54	07E02	14.67	82.28	4.39	3	0.1	0.08	28.1	19344
55	07E01	4.92	82.28	4.24	3	0.1	0.08	28.1	19359
56	05W01	-4.84	57.93	3.89	4	0.1	0.08	24.0	19462
57	05W02	-14.60	57.93	3.76	4	0.1	0.00	24.0	19327
58	05W03	-24.36	57.91	3.61	4	0.1	0.08	24.0	19361
59	05W04	-34.11	57.92	3.44	4	0.1	0.08	24.0	19586
60	07W01	-4.86	82.31	4.07	4	0.1	0.08	28.1	19863
61	07W02	-14.62	82.30	3.89	4	0.1	0.08	28.1	19337
62	07W04	-34.13	82.28	3.56	4	0.1	0.00	28.1	19339
63	07W05	-43.86	82.31	3.32	4	0.1	0.08	28.1	19552

Table A.3: Heliostat positions and channel-specific `Sandfield` parameters (*part* II, *from left to right*): STACEE channel number, heliostat ID, heliostat position (x, y, z), secondary mirror number (*Sec.*), Gaussian facet pointing error (σ_f), canting parameter in units of [km^{-1}] (*Cant.*), opening half-angle of DTIRC (ξ), serial number of PMT (*PMT No.*).

A.3 Electronics simulation parameters

The `Elec` parameter file contains all the user-tunable input parameters necessary to steer the STACEE electronics simulation. An example of a parameter file with the choice of parameters used for all relevant simulations of this work is given in Table A.4.

```
Elec Version                  = 2.5       // version number
Number of PMTs (48/64)        = 64        // - 1 - (explanations below)
Coincidence Mode (1-5)        = 1         // - 2 -
Phillips (0) or UCD (1) Amps  = 0         // - 3 -
Amp saturation on(1)/off(0)?  = 1         // - 4 -
Scrambled mode runlength [s]  = 5.0       // - 5 -
Number of scrambled events    = 0         // - 6 -
Number of laser shots         = 1800      // - 7 -
Filter wheel on(1)/off(0)?    = 1         // - 8 -
Laser time spread sigma [ns]  = 2.5       // - 9 -
Write out triggered events?   = 1         // - 10 -
Write out waveforms?          = 1         // - 11 -
Calculate FADC pedestals?     = 1         // - 12 -
Calculate PMT rates?          = 1         // - 13 -
Single PE pulse shape (0/1)   = 0         // - 14 -
Fake run number               = 99000     // - 15 -
Random seed (0 to use time)   = 0         // - 16 -
Text output on(1)/off(0)?     = 1         // - 17 -
Verbose output?               = 0         // - 18 -
Include Sky Noise?            = 1         // - 19 -
Include RF Noise?             = 1         // - 20 -
RF noise coherent (no is 0)?  = 0         // - 21 -
Discr. Threshold E,W,N [mV]   = -130      // - 22 -
Discr. Threshold S     [mV]   = -130      // - 23 -
Fan-Out DC Offset [mV]        = 0         // - 24 -
L1 coincidence (LTC)          = 5         // - 25 -
L2 coincidence (GTC)          = 5         // - 26 -
McGill Mode (narrow is 0)     = 0         // - 27 -
Narrow mode L1 gate [ns]      = 12        // - 28 -
Discriminator deadtime [ns]   = 7         // - 29 -
TDC deadtime [ns]             = 9         // - 30 -
Time above threshold [ns]     = 0.8       // - 31 -
Pulse shape ratio [ns]        = 4.0       // - 32 -
Transit time spread RMS [ns]  = 0.5       // - 33 -
Relative gain spread          = 0.0       // - 34 -
Canting Parameter [1/km]      = 0.08      // - 35 -
Source zenith angle [deg]     = -1.0      // - 36 -
Source azimuth angle [deg]    = -1.0      // - 37 -
Chan Current Gain Thresh Effic. Refl. Trim  dTrim Laser L.Max  RF-Ampl.
  0   27.25  1.2   120   1.00   0.89   0.0   0.0   1.0   80.0   1.5
  1   25.73  1.2   117   1.00   0.89   0.0   0.0   1.0   80.0   1.5
  2   23.39  1.2   123   1.00   0.89   0.0   0.0   1.0   80.0   1.5
  3   25.52  1.2   117   1.00   0.89   0.0   0.0   1.0   80.0   1.5
                                 ...
 63   33.72  1.2   144   1.00   0.85   3.0   0.0   1.0   80.0   1.5
```

Table A.4: Typical `Elec` input parameter file with default values; see next page for explanations of the individual parameters, listed by line number (comments on the right).

Explanation of the global parameters, by line number:

1. Simulate STACEE-48 or -64 ?
2. Switch between Phillips (pre-2004/05 observing season) and UCD amplifiers.
3. Option to simulate saturation of large pulses, caused by amplifiers.
4. Coincidence mode: IN-TIME (1), SCRAMBLED (2, 5), LASER PHOTODIODE (3) and LASER FLASHER (4) modes are supported.
5. Termination criterion for running on noise with randomised delays: elapsed simulated running time.
6. Termination criterion for running on noise with randomised delays: number of triggers.
7. Number of laser events in LASER mode.
8. Simulates the continuous filter wheel: laser pulse amplitudes increase with time.
9. Laser time spread in nanoseconds, determining the intrinsic timing resolution of the laser.
10. Option to not write a `stoff` raw data output file.
11. Option to not write a merged `stoff` output file.
12. Switch for trace pedestal calculation.
13. Switch for PMT rate calculation.
14. Switch between two types of single PE pulse shape.
15. Run number of the fate data file.
16. Option to use a fixed random seed at run start.
17. Option to toggle text output on/off.
18. Option to toggle verbose text output on/off.
19. Toggle night-sky background (NSB) photons on/off.
20. Toggle 100 MHz RF noise on/off.
21. Toggle coherence of RF noise on/off.
22. Set discriminator thresholds [mV] for 260-level cameras.
23. Set discriminator thresholds [mV] for 220-level cameras.
24. DC offset on Fan-Outs.
25. Cluster trigger condition (LTC).

26. Global trigger condition (GTC).
27. Switch McGill trigger between *narrow* and *wide* coincidence modes.
28. Narrow mode coincidence gate width [ns].
29. Deadtime of the PMT pulse discriminators [ns].
30. The minimum required time above threshold of a PMT pulse in order to fire the discriminator.
31. Integral-to-peak ratio of single PE pulses (measure of pulse width).
32. Transit time jitter RMS of individual PE's in the PMT.
33. Fraction of the nominal gains by which they are smeared out around the latter.
34. Canting parameter (= inverse heliostat canting height ASL) in [1/km].
35. Source zenith angle. (If negative, headings from the shower header are used.)
36. Source azimuth angle (see above).

Explanation of the channel-by-channel parameters:

- The PMT night-sky **currents** are the real PMT anode currents in units of microampere (μA).
- The **gains** are the nominal PMT gains (in units of 10^5) for that channel.
- The '`Thresh`' (**threshold**) column contains the channel specific (*effective*) thresholds.
- The **efficiencies** allow the user to manipulate the total electro-optical throughput of the experiment and/or individual channels.
- The '`Refl.`' (**reflectivities**) column contains the average reflectivities of the secondary mirrors.
- The **trims** are needed to correct for differences in signal propagation time between individual channels.
- **Trim errors** ('`dTrim`') account for the fact that the trims in the actual experiment are never perfect. The numbers given entered are *added* to the signal photon arrival times.
- The '`Laser`' column contains the **laser amplitudes** for each channel, in terms of the *mean* number of photoelectrons released from the PMT photocathode.
- The '`L.Max`' column has only an effect if the filter-wheel option (line -6-) is on, and contains the maximum mean laser amplitudes for each channel.
- These are the RMS values of the modulated 100 MHz **RF noise**, in units of [mV].

APPENDIX B

THE H 1426+428 DATA SET

B.1 Pair-wise statistics

The amount of live time remaining after time cuts, as well as the event numbers and ON−OFF significances before and after padding/grid cuts are listed for each pair in Tables B.1 (2003 data) and B.2/B.3 (2004 data). The *padding cut* is a static gate (16 ns width) cut after software padding, and for the *grid cut* it was required that $\xi < 0.37$ after applying a correction for the South camera fraction.

Run numbers		Live times		Time cuts only			Padding and Grid cut		
ON	OFF	$t_{\mathrm{ON}}\,[\mathrm{s}]$	$t_{\mathrm{OFF}}\,[\mathrm{s}]$	N_{ON}	N_{OFF}	$S_0\,[\sigma]$	N_{ON}	N_{OFF}	$S\,[\sigma]$
11088	11089	1153.43	1150.69	6192	6431	-2.26	1251	1234	0.28
11106	11107	766.23	766.88	3999	3909	1.05	602	605	-0.07
11121	11122	1509.74	1511.41	7775	7796	-0.10	1593	1539	1.00
11123	11124	689.02	689.26	3574	3521	0.64	557	571	-0.41
11504	11505	1436.11	1437.11	7125	7176	-0.38	1129	1102	0.59
11525	11526	1436.02	1437.51	7232	7120	1.00	1166	1180	-0.26
11550	11551	1303.97	1307.84	6662	6568	0.99	1086	1113	-0.51
11552	11553	1510.55	1513.00	7313	7214	0.92	1134	1044	1.97
11578	11579	1306.78	1311.25	6775	6695	0.89	1102	1121	-0.32
11580	11581	1428.93	1431.38	6716	6673	0.47	976	972	0.13
11598	11599	1381.04	1381.06	6634	6804	-1.47	1272	1322	-0.98
11600	11601	1245.65	1247.88	6222	5939	2.67	1018	953	1.50
11644	11645	1360.86	1361.35	9382	9395	-0.07	1589	1607	-0.31
11646	11647	1465.43	1466.11	9938	10016	-0.52	1569	1526	0.79
11677	11678	1545.15	1546.92	7137	7120	0.21	920	914	0.16
11700	11701	1536.11	1536.39	7965	7929	0.30	1337	1313	0.47
11702	11703	1282.26	1279.41	6059	6087	-0.38	747	685	1.60
11782	11783	400.41	401.14	1858	1838	0.38	292	301	-0.35
11840	11841	965.57	968.68	4631	4553	0.97	697	707	-0.21
11843	11844	1431.35	1435.97	6303	6043	2.52	824	777	1.24
11915	11916	1471.31	1473.96	5148	5190	-0.32	290	323	-1.31
11978	11979	1655.30	1655.33	7557	7495	0.51	1130	1072	1.24
12046	12047	1413.71	1416.01	6621	6559	0.63	1063	1082	-0.37
12049	12050	1412.58	1414.66	5896	5631	2.55	731	625	2.91
12114	12115	869.80	871.09	3822	3784	0.50	405	474	-2.31

Table B.1: Pair-wise statistics for 2003's observations of H 1426+428.

Run numbers		Live times		Time cuts only			Padding and Grid cut		
ON	OFF	$t_{\mathrm{ON}}\,[\mathrm{s}]$	$t_{\mathrm{OFF}}\,[\mathrm{s}]$	N_{ON}	N_{OFF}	$S_0\,[\sigma]$	N_{ON}	N_{OFF}	$S\,[\sigma]$
13953	13952	1496.37	1495.19	7719	7762	-0.39	933	869	1.49
14008	14009	349.18	347.69	2078	2187	-1.81	352	376	-0.95
14036	14035	1638.75	1638.66	9746	9872	-0.90	1161	1199	-0.78
14038	14037	0.00	0.05	0	0	0.00	0	0	0.00
14039	14040	85.06	84.45	499	541	-1.42	84	95	-0.87
14063	14064	711.44	713.92	3836	3767	0.94	475	496	-0.62
14092	14091	1382.16	1381.50	6433	6481	-0.45	726	732	-0.17
14094	14093	1558.75	1557.85	7542	7571	-0.27	1238	1270	-0.65
14095	14096	159.58	159.97	786	763	0.63	123	130	-0.42
14123	14122	1544.38	1541.88	7633	7747	-1.02	1087	1053	0.70
14126	14127	1535.21	1535.51	7732	7749	-0.12	1336	1315	0.41
14128	14129	1461.01	1461.66	7396	7352	0.39	950	969	-0.42
14147	14146	1322.79	1323.39	6279	6327	-0.40	1021	1055	-0.74
14150	14151	0.00	0.00	0	0	0.00	0	0	0.00
14186	14187	688.51	690.52	4253	4099	1.82	626	609	0.54
14247	14246	1445.32	1445.24	7059	6955	0.87	1149	1070	1.68
14248	14249	1520.98	1522.29	7286	7288	0.04	1049	1025	0.55
14265	14266	1414.87	1415.07	8625	8696	-0.53	1186	1238	-1.05
14280	14281	1493.34	1496.91	8464	8335	1.15	1074	1066	0.23
14282	14283	567.47	565.74	3074	3169	-1.32	376	391	-0.58
14305	14304	1513.55	1515.11	7478	7295	1.57	1312	1278	0.69
14306	14307	1457.19	1458.72	7138	7109	0.31	988	1037	-1.07
14308	14309	1456.50	1455.90	6902	7008	-0.92	844	867	-0.56
14324	14323	566.65	565.23	3046	3149	-1.41	352	348	0.12
14345	14344	1185.93	1186.15	5647	5568	0.76	906	828	1.88
14346	14347	1524.69	1528.48	7226	7013	1.93	1093	1029	1.45
14348	14349	1503.13	1501.81	6866	6956	-0.82	820	819	0.01
14367	14366	1311.76	1311.66	8633	8567	0.50	1479	1419	1.11
14368	14369	1306.21	1305.09	8377	8387	-0.13	1132	1134	-0.06
14388	14389	869.46	869.53	5602	5527	0.72	904	965	-1.41
14391	14392	1280.23	1282.17	9205	9161	0.43	1059	1056	0.10
14395	14396	1434.66	1435.34	7891	7733	1.29	861	837	0.59
14411	14412	1410.75	1415.49	7796	7645	1.42	1216	1242	-0.44
14430	14431	1386.45	1389.12	7794	7848	-0.31	1205	1134	1.51
14433	14434	1139.37	1140.49	6330	6342	-0.05	685	689	-0.09
14437	14438	1389.79	1390.59	7284	7293	-0.04	617	624	-0.19
14448	14449	1225.85	1229.10	7417	7378	0.48	863	906	-0.97
14451	14452	621.99	624.64	3679	3568	1.48	278	298	-0.78
14592	14593	1501.74	1504.93	8856	8963	-0.66	1101	1216	-2.34
14595	14596	1250.31	1248.23	7116	7203	-0.83	788	794	-0.18
14598	14599	0.00	0.00	0	0	0.00	0	0	0.00
14598	14600	1183.23	1183.02	6234	6320	-0.78	472	480	-0.26

Table B.2: Pair-wise statistics for 2004's observations of H 1426+428 (*part I*).

Run numbers		Live times		Time cuts only			Padding and Grid cut		
ON	OFF	$t_{\mathrm{ON}}\,[\mathrm{s}]$	$t_{\mathrm{OFF}}\,[\mathrm{s}]$	N_{ON}	N_{OFF}	$S_0\,[\sigma]$	N_{ON}	N_{OFF}	$S\,[\sigma]$
14614	14613	1324.04	1320.32	7253	7184	0.40	1016	961	1.17
14615	14616	1116.62	1114.06	5963	6053	-0.95	958	992	-0.82
14619	14620	1340.94	1344.06	7097	7115	-0.01	841	881	-0.92
14621	14622	738.23	740.30	3836	3784	0.72	395	403	-0.24
14643	14644	0.18	0.00	0	0	0.00	0	0	0.00
14661	14662	1404.60	1405.21	7937	7953	-0.10	1386	1328	1.12
14665	14666	482.30	481.25	2821	2834	-0.25	331	353	-0.87
14678	14677	1080.11	1078.28	6915	6894	0.08	864	783	1.96
14697	14698	1409.14	1412.01	5907	5928	-0.08	800	799	0.07
14699	14700	1423.65	1422.01	5447	5584	-1.36	589	598	-0.28
14728	14729	824.57	826.99	4853	4874	-0.07	800	811	-0.22
14731	14732	772.71	777.12	4468	4423	0.75	549	533	0.58
14734	14735	0.00	0.00	0	0	0.00	0	0	0.00
14734	14736	236.55	237.27	1319	1274	0.96	112	132	-1.26
14766	14767	1479.87	1486.85	8700	8645	0.73	1163	1081	1.84
14769	14770	1480.40	1485.01	8489	8216	2.31	907	921	-0.26
14771	14772	26.58	26.72	143	136	0.46	11	13	-0.40
14797	14798	1044.83	1048.56	5535	5496	0.56	620	669	-1.30
14800	14801	1354.97	1356.81	6762	6743	0.24	630	685	-1.49
14838	14837	803.28	803.06	3607	3593	0.15	460	435	0.83
14840	14841	726.22	727.66	3103	3113	-0.05	326	371	-1.68
14843	14844	984.27	987.33	3844	3761	1.09	247	233	0.67
14860	14861	1294.16	1297.85	6686	6538	1.45	850	814	0.94
14863	14864	1369.65	1371.02	6319	6239	0.77	576	531	1.37
14879	14880	1177.46	1177.51	6212	6287	-0.67	741	780	-1.00
14882	14883	0.00	0.00	0	0	0.00	0	0	0.00

Table B.3: Pair-wise statistics for 2004's observations of H 1426+428 (*part II*). N_{ON} and N_{OFF} are the event numbers in the ON and OFF run, respectively, passing the cuts indicated; S_0 is the raw significance and S the final significance after all cuts.

REFERENCES

Abraham, R.G., Crawford, C.S. & McHardy, I.M., *Optical imaging of BL Lac host galaxies*, MNRAS **252**, 482-504 (1991).

Abraham, J. *et al.* (Pierre Auger Collaboration), *Anisotropy studies around the galactic centre at EeV energies with the Auger Observatory*, Astropart. Phys. **27** (4), 244-253 (2006). [arXiv:astro-ph/0607382]

Aharonian, F. *et al.*, *TeV gamma rays from the blazar H 1426+428 and the diffuse extragalactic background radiation*, A&A **384**, L23-L26 (2002). [arXiv:astro-ph/0202072]

Aharonian, F. *et al.*, *Observations of H1426+428 with HEGRA. Observations in 2002 and reanalysis of 1999&2000 data*, A&A **403**, 523-528 (2003). [arXiv:astro-ph/0301437]

Aharonian, F. *et al.*, *Discovery of very-high energy γ-rays from the Galactic Centre ridge*, Nature **439**, 695-698 (2006).

Aharonian, F. *et al.*, *A low level of extragalactic background light as revealed by gamma-rays from blazars*, Nature **440** (7087), 1018-1021 (2006). [arXiv:astro-ph/0508073]

Aharonian, F. *et al.*, *Discovery of Very High Energy γ-Ray Emission from the BL Lac Object H 2356-309 with the H.E.S.S. Cherenkov Telescopes*, A&A **455**, 461-466 (2006). [arXiv:astro-ph/0607569]

Aharonian, F. *et al.*, *Fast variability of TeV γ-rays from the radio galaxy M 87*, Science **314**, 1424-1427 (2006). [arXiv:astro-ph/0612016]

Aharonian, F. *et al.*, *Discovery of VHE γ-rays from the distant BL Lac 1ES 0347-121*, A&A **473**, L25-L28 (2007).

Aharonian, F. *et al.*, *An exceptional VHE gamma-ray flare of PKS 2155-304*, ApJ **664**, L71 (2007). [arXiv:0706.0797]

Aharonian, F. *et al.*, *New constraints on the Mid-IR EBL from the HESS discovery of VHE gamma rays from 1ES 0229+200*, A&A **475**, L9-L13 (2007). [arXiv:0709.4584]

Aharonian, F. *et al.*, *Detection of VHE gamma-ray emission from the distant blazar 1ES 1101-232 with H.E.S.S. and broadband characterisation*, A&A **470**, 475-489 (2007). [arXiv:0705.2946]

Akhperjanian, A. & Sahakian, V., *Performance of a 20 m diameter Cherenkov imaging telescope*, Astropart. Phys. **21** 2, 149-161 (2004).

Albert, J. *et al.*, *Discovery of Very High Energy Gamma Rays from 1ES 1218+30.4*, ApJ **642**, L119-L122 (2006). [arXiv:astro-ph/0603529]

Albert, J. *et al.*, *Discovery of Very High Energy γ-Ray Emission from the Low-Frequency-peaked BL Lacertae Object BL Lacertae*, ApJ **666**, L17-L20 (2007). [arXiv:astro-ph/0703084v2]

Albert, J. *et al.*, *Discovery of Very High Energy γ-Rays from 1ES 1011+496 at z = 0.212*, ApJ **667** (1), L21-L24 (2007). [arXiv:0706:4435v1]

Albuquerque, I. F. M. & Smoot, G. F., *GZK cutoff distortion due to the energy error distribution shape*, Astropart. Phys. **25**, 375-379 (2006). [arXiv:astro-ph/0504088v3]

Amenomori, M. *et al.*, *A Northern Sky Survey for Steady TeV Gamma-Ray Point Sources Using the Tibet Air Shower Array*, ApJ **633**, 1005-1012 (2005). [arXiv:astro-ph/0502039]

ANTARES Collaboration, *Transmission of light in deep sea water at the site of the Antares neutrino telescope*, Astropart. Phys. **23**, 131-155 (2005). [arXiv:astro-ph/0412126]

Appenzeller, I. *et al.*, *Successful commissioning of FORS1 - the First Optical Instrument on the VLT*, The Messenger **94**, 1-6 (1998).

ATIC Collaboration, *Cosmic Rays at 120,000 feet above Antarctica - The Advanced Thin Ionization Calorimeter (ATIC) Experiment*, Inv. talk, Crary Lab Wednesday Night Science Lecture, McMurdo, Antarctica (2005).

Atkins, R. *et al.*, *TeV Gamma-Ray Survey of the Northern Hemisphere Sky Using the Milagro Observatory*, ApJ **608**, 680 (2004). [arXiv:astro-ph/0403097]

Auger Collaboration, *Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei*, Astropart. Phys. **29**, 188 (2008). [arXiv:0712.2843]

Bai, J. M. & Lee, M. G., *New Evidence for the Unified Scheme of BL Lacertae Objects and FR I Radio Galaxies*, ApJ **548** 1, 244-248 (2001). [arXiv:astro-ph/0012340]

Baixeras, C. *et al.*, *MAGIC II*, Proc. 29th ICRC (Pune, India), **5**, 227 (2005). [arXiv:astro-ph/0508274]

Balbus, S. A. & Hawley, J. F., *A powerful local shear instability in weakly magnetized disks. IV. Nonaxisymmetric perturbations*, ApJ **400**, 610 (1992).

Barnett, R. M. *et al.*, *Review of Particle Physics*, Phys. Rev. D **54** (1996).

Barrau, A. *et al.*, *The CAT Imaging Telescope for Very-High-Energy Gamma-Ray Astronomy*, NIM A **416**, 278 (1998).

Bekenstein, J. D., *Black Holes: Physics and Astrophysics*, "Neutrinos and Explosive Events in the Universe" (ISBN 1-4020-3746-5), **209**, 149 (2005). [arXiv:astro-ph/0407560]

Berezinsky, V., Gasisov, A. Z. & Grigorieva, S. I., *Dip in UHECR spectrum as signature of proton interaction with CMB*, Phys. Lett. B **612**, 147-153 (2005). [arXiv:astro-ph/0502550]

Berezinsky, V., *Dip in UHECR and Transition from Galactic to Extragalactic Cosmic Rays*, Journal of Physics: Conf. Ser. **47** (1), 142-153 (2006). [arXiv:astro-ph/0509069]

Berge, D., *Development of an Algorithm for the Shower Reconstruction with the H.E.S.S. Telescope System*, Diplomarbeit, Humboldt-Universität zu Berlin (2002).

Bergman, D. R. *Observation of the GZK cutoff using the HiRes Detector* Nucl. Phys. Proc. Suppl. **165**, 19-26 (2007). [arXiv:astro-ph/0609453]

Bernlöhr, K., *Impact of atmospheric parameters on the atmospheric Cherenkov technique*, Astropart. Phys. **12**, 255-268 (2000).

Bignami, G. F. *et al.*, *The COS-B experiment for gamma-ray astronomy*, Space Sci. Instr. **1**, 245-268 (1975).

Blandford, R. D. & Znajek, R. L., *Electromagnetic extraction of energy from Kerr black holes*, MNRAS **179**, 433-456 (1977).

Blandford, R. D. & Rees, M. J., *Some comments on radiation mechanisms in Lacertids*, Proc. Pittsburgh Conf. on BL Lacs, 328-341 (1978).

Blandford, R. D. & Königl, A., *Relativistic jets as compact radio sources*, ApJ **232**, 34-48 (1979).

Blandford, R. D. & Payne, D. G., *Hydromagnetic flows from accretion discs and the production of radio jets*, MNRAS **199**, 883-903 (1982).

Blandford, R. D. & Eichler, D., *Particle Acceleration at Astrophysical Shocks - a Theory of Cosmic-Ray Origin*, Phys. Rep. **145**, 1 (1987).

Blandford, R. D. & Levinson, A., *Pair cascades in extragalactic jets. 1: Gamma rays*, ApJ **441**, 79-95 (1995).

Blasi, P. & Vietri, M., *On Particle Acceleration around Shocks. II. A Fully General Method for Arbitrary Shock Velocities and Scattering Media* ApJ **626**, 877-886 (2005).

Blazejowski, M. *et al.*, *A Multi-wavelength view of the TeV blazar Markarian 421: Correlated variability, flaring, and spectral evolution.*, ApJ **630**, 130-141 (2005). [arXiv:astro-ph/0505325]

Blustin, A. J., Page, M. J. & Branduardi-Raymont, G., *Intrinsic absorbers in BL Lac objects: The XMM-Newton view* A&A **417**, 61-70 (2004).

Boone, L. M., *Introduction to the* `Hermes` *Electronics Simulator*, STACEE Internal Collaboration Note 00-07 (2000).

Boone, L. M., *Observations of Markarian 421 with the Solar Tower Atmospheric Cherenkov Effect Experiment*, Ph.D. thesis, University of California, Santa Cruz (2002).

Boone, L. M. *et al.*, *STACEE Observations of Markarian 421 during an Extended Gamma-Ray Outburst*, ApJL **579**, L5-L8 (2002).

Borione, A. *et al.*, *A large air shower array to search for astrophysical sources emitting γ-rays with energies >=1014 eV*, NIM A **346**, 329-352 (1994).

Böttcher, M. & Dermer, C. D., *An Evolutionary Scenario for Blazar Unification*, ApJ **564** (1), 86-91 (2002). [arXiv:astro-ph/0106395v2]

Böttcher, M., Mukherjee, R. & Reimer, A., *Predictions of the High-Energy Emission from BL Lacertae Objects: The Case of W Comae*, ApJ **581**, 143-154 (2002). [arXiv:astro-ph/0208171]

Böttcher, M. & Reimer, A., *Modeling the multiwavelength spectra and variability of BL Lacertae in 2000* ApJ **609**, 576-588 (2004).

Bramel, D., *Pass1 Analysis Chain with* `stoff_pair`, STACEE Internal Collaboration Note 03-06 (2003).

Bramel, D., *Observations of the BL Lac Object 3C66A with the Solar Tower Atmospheric Cherenkov Effect Experiment*, Ph.D thesis, Columbia University (2005).

Carilli, C. L. & Barthel, P. D., *Cygnus A*, A&A Rev. **7**, 1-54 (1996).

Carrol, B. W. & Ostlie, D. A., *An introduction to modern astrophysics*, Addison-Wesley (1996).

Carson, J. E., Kildea, J., Ong, R. A. *et al.*, *The Energy Spectrum of the Blazar Markarian 421 Above 130 GeV*, ApJ **662** (1), 199-204 (2007). [arXiv:astro-ph/0612562]

Carson, J. E., *GLAST: physics goals and instrument status*, Proc. TeV Part. Astroph. II, Madison, WI (2006). [arXiv:astro-ph/0610960]

Cavaliere, A. & D'Elia, V., *The Blazar Main Sequence*, ApJ **571**, 226-233 (2002).

Chang, J. *et al.*, *The Electron Spectrum above 20 GeV Measured by ATIC-2*, Proc. 29th ICRC (Pune, India), **3**, 1-4 (2005).

Costamante, L., Ghisellini, G. *et al.*, *Looking for high energy peaked blazars*, Mem. della Soc. Astr. Italiana **72** (1), 153-155 (2001). [arXiv:astro-ph/0007020]

Costamante, L. *et al.*, *Extreme Synchrotron BL Lac Objects*, A&A **371**, 512-526 (2001). [arXiv:astro-ph/0103343]

Costamante, L. *et al.*, *The SED of the TeV BL Lac 1ES 1426+428 after correction for the TeV-IR absorption*, New Astron. Rev. **47**, 677-680 (2003). [arXiv:astro-ph/0301211]

Costamante, L. *et al.*, *On the intrinsic TeV spectra of 1ES 1426+428 and 1ES 1959+650*, Proc. Part. Accel. in Astr. Obj., Crakow, Poland (2003).

Costamante, L. *et al.*, *Constraining the Cosmic Background Light with four BL Lac TeV spectra*, New Astron. Rev. **48**, 469-472 (2004). [arXiv:astro-ph/0308025]

Costamante, L., *A low density of the Extragalactic Background Light revealed by the H.E.S.S. spectra of the BLLac objects 1ES 1101-232 and H 2356-309*, Astrophys. & Space Sci. **309**, 487-495 (2007). [arXiv:astro-ph/0612709]

Covault, C. E., Oser, S. M. & Scalzo, R. A., *Sewer Pipe optical Tests on PMTs and DTIRCs*, STACEE Internal Collaboration Note 98-08 (1998).

Covault, C., *The stacee_banks software package to treat STACEE data*, STACEE Internal Collaboration Note 00-05 (2000).

Dai, Z. G., Zhang, B. *et al.*, *GeV Emission from TeV Blazars and Intergalactic Magnetic Fields*, ApJ **580** (1), L7-L10 (2002). [arXiv:astro-ph/0209091]

Dar, A., *Fireball and Cannonball Models of Gamma-Ray Bursts Confront Observations*, Conf. Proc. 2005 Frascati Workshop, Vulcano, Italy (2005). [arXiv:astro-ph/0511622]

Daum, A. *et al.*, *First results on the performance of the HEGRA IACT array*, Astropart. Phys. **8** (1-2), 1-11 (1997).

Davies, J. & Cotton, E., *Design of the quartermaster solar furnace*, E. Journal of Solar Energy **1**, 2-3, 16-22 (1957).

d'Avezac, P., *Contribution à la calométrie du télescope spatial à rayon γ GLAST, et étude des cascades électron-photon sur le fond diffus extragalactique*, Ph.D thesis, Ecole Polytechnique, Paris (2006).

d'Avezac, P., Dubus, G. & Giebels, B., *Cascading on extragalactic background light*, A&A **469**, 857-860 (2007). [arXiv:0704.3910v1]

Della Valle, M., *Supernovae shedding light on Gamma-Ray Bursts*, Nuovo Cimento **28C**, 563-573 (2005).

de Naurois, M., *Étude de la photocathode*, CELESTE Internal Collaboration Note (1998).

de Naurois, M., *L'expérience CELESTE: Reconversion d'une centrale solaire pour l'astronomie γ: Première observation de la Nébuleuse du Crabe et du Blazar Markarian 421 entre 30 et 300 GeV.*, Ph.D. Thesis, Laboratoire Leprince-Ringuet (2000).

de Naurois, M., Guy, J., Djannati-Ataï, A. & Tavernet, J. P., Proc. 28th ICRC (Tsukuba, Japan), OG 2.5, p. 2907 (2003).

de Naurois, M., *High Energy Gamma Rays*, Proc. XXIV Physics In Collision, Boston, MA (2004).

Dermer, C. D., Schlickeiser, R. & , A., *High-energy gamma radiation from extragalactic radio sources*, A&A **256** (2), L27-L30 (1992).

de Robertis, M. M., Yee, H. K. C. & Hayhoe, K., *A CCD Study of the Environment of Seyfert Galaxies. II. Testing the Interaction Hypothesis*, ApJ **496**, 93 (1998).

Di Girolama *et al.*, *Performance of the Argo-YBJ detector operated in scaler mode*, Proc. 29th ICRC (Pune, India), **4**, 431 (2005).

Djannati-Ataï, A. *et al.*, *Detection of the BL Lac object 1ES 1426+428 in the Very High Energy gamma-ray band by the CAT Telescope from 1998-2000*, A&A **391**, L25-L28 (2002). [arXiv:astro-ph/0207618]

Donato, D., Ghisellini, G., Tagliaferri, G. & Fossati, G., *Hard X-ray properties of blazars*, A&A **375**, 739-751 (2001). [arXiv:astro-ph/0105203]

Dwek, E. & Krennrich, F., *Simultaneous Constraints on the Spectrum of the Extragalactic Background Light and the Intrinsic TeV Spectra of Markarian 421, Markarian 501, and H1426+428*, ApJ **618** (2), 657-674 (2005). [arXiv:astro-ph/0406565]

Dwek, E., Krennrich, F. & Arendt, R. G., *Is there an imprint of primordial stars in the TeV γ-ray spectrum of blazars?*, ApJ **634**, 155-160 (2005). [arXiv:astro-ph/0508133]

Egorova, V. P. *et al.*, *The spectrum features of UHECRs below and surrounding GZK*, Nucl. Phys. B (Proc.Suppl.) **136**, 3-11 (2004). [arXiv:astro-ph/0408493]

Engel, R., Gaisser, T. K., Lipari, P. & Stanev, T., *Air Shower Calculations With the New Version of SIBYLL*, Proc. 26th ICRC (Salt Lake City, USA), **1**, 415 (1999).

ESA/NASA, the AVO project and Paolo Padovani (2006).

Falcone, A. D. *et al.*, *X-Ray Spectral Variability of Extreme BL Lacertae Active Galactic Nucleus H1426+428*, ApJ **601** (1), 165-172 (2004). [arXiv:astro-ph/0310226]

Falomo, R., Scarpa, R., Treves, A. & Urry, C. M., *The Hubble Space Telescope Survey of BL Lacertae Objects. III. Morphological properties of low-redshift host galaxies*, ApJ **542**, 731-739 (2000).

Fan, Y. Z. *et al.*, *Strong GeV emission accompanying TeV blazar H1426+428*, A&A **415**, 483-486 (2004). [arXiv:astro-ph/0310893]

Fanaroff, B. L. & Riley, J. M., *The morphology of extragalactic radio sources of high and low luminosity*, MNRAS **167**, 31-36 (1974).

Fassò, A., Ferrari, A., Ranft, J. & Sala, P. R., *FLUKA: Status and Prospective of Hadronic Applications*, Proc. Monte Carlo 2000 Conf., Lisbon, 2000; A. Kling, F. Barao, M. Nakagawa, P. Vaz *eds.*, Springer (Berlin) 955 (2001).

Feldman, G. J. & Cousins, R. D., *A Unified Approach to the Classical Statistical Analysis of Small Signals*, Phys. Rev. **D 57**, 3873-3889 (1998). [arXiv:physics/9711021v2]

Ferenc, D., *The MAGIC gamma-ray observatory*, NIM A **553**, 274-281 (2005).

Fermi, E., *On the Origin of the Cosmic Radiation*, Phys. Rev. **75**, 1169-1174 (1949).

Ferrari, F. & Szuszkiewicz, E., *Cosmic ray recipes*, arXiv:astro-ph/0601158 (2006).

Fesefeldt, H. *et al.*, *GHEISHA, the simulation of hadronic showers: physics and applications*, Rep. PITHA-85/02, RWTH Aachen (1985).

Fichtel, C. E. *et al.*, *High-Energy Gamma-Ray Results from Second Small Astronomy Satellite*, ApJ **198**, 163 (1975).

Fortin, P., *New results from the McGill Pencil Beam Test Facility*, STACEE Internal Collaboration Note 00-09 (2000).

Fortin, P., *STACEE Gain Measurement Using FGIT*, STACEE Internal Collaboration Note 02-07 (2002).

Fortin, P., *Timing Resolution of STACEE Channels*, STACEE Internal Collaboration Note 03-11 (2003).

Fortin, P., *Observations of the Crab Nebula and pulsar in the optical and γ-ray bands with STACEE*, Ph.D. Thesis, McGill University (2005).

Fossati, G., Maraschi, L. *et al.*, *A unifying view of the spectral energy distributions of blazars*, MNRAS **299**, 433-448 (1998).

Fowler, J. W., *Composition and Spectrum of Cosmic Rays at the Knee Measured by the CASA-BLANCA Experiment*, Ph.D. Thesis, University of Chicago (2000).

Gaisser, T. K. & Stanev, T., *Cosmic Rays*, Euro. Phys. J. **C15**, 150 (2000).

Gallant, Y. A., *Particle Acceleration at Relativistic Shocks*, Relativ. Flows in Astroph. **589**, 24 (2002). [arXiv:astro-ph/0201243]

Gehrels, N. *et al.*, *The Swift Gamma-Ray Burst Mission*, ApJ **611**, 1005-1020 (2004).

Gentile, S., *Detection and measurement of gamma rays with the AMS-02 detector*, Proc. 20th ECRS (Lisbon, Por., 2006) (2007). [arXiv:astro-ph/0702306]

Ghez, A. M., Salim, S. *et al.*, *Stellar orbits around the Galactic Center Black Hole*, ApJ **620**, 744-757 (2005).

Ghisellini, G., Celotti, A., Fossati, G., Maraschi, L. & Comastri, A., *A theoretical unifying scheme for gamma-ray bright blazars*, MNRAS **301**, 451-468 (1998).

Ghisellini, G., Celotti, A. & Costamante, L., *Low power BL Lacertae objects and the blazar sequence*, A&A **386**, 833-842 (2002).

Ghisellini, G., Tavecchio, F. & Chiaberge, M., *Structured jets in TeV BL Lac objects and radiogalaxies. Implications for the observed properties*, A&A **432** (2), 401-410 (2005). [arXiv:astro-ph/0406093]

Giacconi, R. *et al.*, *The Uhuru catalog of X-ray sources*, ApJ **178**, 281-308 (1972).

Giacconi, R. & Gursky, H., *X-Ray Astronomy*, Springer (1974).

Gingrich, D., *The STACEE Ground-Based Gamma-Ray Detector*, IEEE Trans. Nucl. Sci. NS-52 (6) 2977-2985 (2005).

Giommi, P., Massaro, E., Padovani, P., Perri, M. *et al.*, *ROXA J081009.9+384757.0: a 10^{47} erg s^{-1} blazar with hard X-ray synchrotron peak or a new type of radio loud AGN?*, A&A **468**, 97-101 (2007). [arXiv:astro-ph/0703491]

Götting, N., *Nachweis von TeV-γ-Strahlung aus der Richtung der Blazare H 1426+428 und 1ES 1959+650 sowie der Radiogalaxie M 87 mit den HEGRA-Cherenkov-Teleskopen*, Ph.D. thesis, Universität Hamburg (2006).

Gould, R. J. & Schréder, G., *Pair Production in Photon-Photon Collisions*, Phys. Rev. **155**, 1404-1407 (1967).

Greisen, K., *End to the Cosmic-Ray Spectrum?*, Phys. Rev. Lett. **16**, 748 (1966).

Guy, J., Renault, C. *et al.*, *Constraints on the Cosmic Infra-Red Background based on BeppoSAX and CAT spectra of Mrk501*, A&A **359**, 419-428 (2000). [arXiv:astro-ph/0004355]

Haino, S. *et al.*, *Measurements of primary and atmospheric cosmic-ray spectra with the BESS-TeV spectrometer*, Phys. Lett. B **594** (1-2), 35-46 (2004).

Hanna, D. S. & Mukherjee, R., *The laser calibration system for the STACEE ground-based gamma ray detector*, NIM A **482**, 271-280 (2002).

Hanna, D. S. *et al.*, *The STACEE-32 ground based gamma-ray detector*, NIM A **491**, 126-151 (2002).

Hanna, D. S., *Gain Drifts Seen With Laser Runs*, STACEE Internal Collaboration Note 02-02 (2002).

Hanna, D., *A New Way to Monitor Phototube Gains in STACEE*, STACEE Internal Collaboration Note 03-04 (2003).

Hartman, R. C. *et al.*, *The Third EGRET Catalog of High-Energy Gamma-Ray Sources*, ApJS **123**, 79-202 (1999).

Hartman, R. C. *et al.*, *Multiepoch multiwavelength spectra and models for blazar 3C 279*, ApJ **553**, 683-694 (2001).

Hauschildt, T. & Steele, D., *Latest results of AMANDA*, Proc. 28th ICRC (Tsukuba, Japan), HE 2.3, 1305 (2003).

Hauser, M. G. & Dwek, E., *The Cosmic Infrared Background: Measurements and Implications*, Ann. Rev. of Astron. & Astroph. **39**, 249-307 (2001). [arXiv:astro-ph/0105539]

Hayashida, N. *et al.*, *Possible Clustering of the Most Energetic Cosmic Rays within a Limited Space Angle Observed by the Akeno Giant Air Shower Array*, Phys. Rev. Lett. **77**, 1000 (1996).

Heck, D., Knapp, J., Capdevielle, J. N., Schatz, G. & Thouw, T., *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*, Rep. FZKA 6019, Forschungszentrum Karlsruhe (1998).

Heck, D., Risse, M. & Knapp, J., *Comparison of Hadronic Interaction Models at Auger Energies*, Nucl. Phys. B (Proc. Suppl.) **122**, 364-367 (2003).

Heck, D., *Low-Energy Hadronic Interaction Models*, Proc. XIII ISVHECRI, Pylos, Greece (2004). [arXiv:astro-ph/0410735]

HEGRA collaboration, *HEGRA Contributions to the 28th International Cosmic Ray Conference*, Proc. 28th ICRC, Tsukuba, Japan (2003). [arXiv:astro-ph/0307334]

Heitler, M., *The Quantum Theory of Radiation*, Oxford, Clarendon (1960).

Herterich, K., *Absorption of gamma rays in intense X-ray sources*, Nature **250**, 311-313 (1974).

Hess, V. F., *Observation of penetrating radiation in seven free balloon flights*, Phys. Zeitschr. **13**, 1084 (1912).

Heywood, I. *et al.*, *The prevalence of FR I radio quasars*, MNRAS , *in press* (2007). [arXiv:0708.1145]

Hillas, A. M., *Cherenkov light images of EAS produced by primary gamma*, Proc. 19th ICRC (La Jolla, USA), **3**, 445-448 (1985).

Hillas, A. M., *The MOCCA Program: Monte-Carlo Cascades*, Proc. 24th ICRC (Rome, Italy), **1**, 270 (1995).

Hillas, A. M. *et al.*, *The Spectrum of TeV Gamma Rays from the Crab Nebula*, ApJ **503**, 744 (1998).

Hinton, J. A., *A comparison of the MOCCA and CORSIKA air-shower simulation codes*, STACEE Internal Collaboration Note 00-10 (2000).

Hinton, J. A., *Sewer pipe tests of the new DTIRC/PMT gluing scheme*, STACEE Internal Collaboration Note 00-11 (2000).

Hinton, J. A., private communication (2001).

Hinton, J. A., Dib, R. & Scalzo, R. A., *Measurements of Heliostat Bias Drift and Light Collection Using CCD Sunspot Images*, STACEE Internal Collaboration Note 00-14 (2000).

Hinton, J. A., private communication (2002).

Hinton, J. A., *The status of the H.E.S.S. project*, New Astron. Rev. **48**, 331-337 (2004). [arXiv:astro-ph/0403052]

Hoffman, C. M., Sinnis, G., Fleury, P. & Punch, M., *Gamma-ray astronomy at high energies* Rev. Mod. Phys. **71**, 897-936 (1999).

Holder, J. *et al.*, *The first VERITAS telescope*, Astropart. Phys. **25** 6, 391-401 (2006). [arXiv:astro-ph/0604119]

Horan, D. *et al.*, *Detection of the BL Lacertae Object H1426+428 at TeV Gamma-Ray Energies*, ApJ **571** 2, 753-762 (2002). [arXiv:astro-ph/0202185]

Horneffer, A. *et al.*, *LOPES Detecting Radio Emission from Cosmic Ray Air Showers*, Proc. SPIE (2004). [arXiv:astro-ph/0409641]

Horns, D. *et al.*, *High Energy Emission from H1426+428 and Absorption on the Extragalactic Background Light*, Proc. 28th ICRC, OG 2.3, 2651 (2003).

Horns, D. *et al.*, *TeV observations of H1426+428 with HEGRA*, New Astron. Rev. **48** (5-6), 387-390 (2004).

IceCube Collaboration, *First Year Performance of The IceCube Neutrino Telescope*, Astropart. Phys. **26** (3), 155-173 (2006). [arXiv:astro-ph/0604450]

Ichimaru, S. *Bimodal behavior of accretion disks - Theory and application to Cygnus X-1 transitions*, ApJ **214**, 840-855 (1977).

Itoh, C. *et al.*, *Erratum: Detection of diffuse TeV gamma-ray emission from the nearby starburst galaxy NGC 253*, Astron. Astrophys., in press (2006).

Jackson, J. D., *Classical Electrodynamics*, 2nd ed., John Wiley & Sons, 1975.

Januzzi, B. T., Smith, P. S. & Elston, R., *Optical polarimetry and photometry of X-ray selected BL Lacertae objects*, ApJS **85** 2, 265-291 (1992).

Jones, T. W., O'dell, S. L. & Stein, W. A., *Physics of Compact Nonthermal Sources. Theory of Radiation Processes*, ApJ **188**, 353-368 (1974).

Kalmykov, N. N., Ostapchenko, S. S. & Pavlov, A. I., *Quark-Gluon-String Model and EAS Simulation Problems at Ultra-High Energies*, Nucl. Phys. B (Proc. Suppl.) **52B** (4), 17-28 (1997).

Kampert, K.-H. *et al.*, *Cosmic Ray Energy Spectra and Mass Composition at the Knee - Recent Results from KASKADE*, Nucl. Phys. B **136**, 273-281 (2004). [arXiv:astro-ph/0410559]

Katarzyński, K., Ghisellini, G., Tavecchio, F., Gracia, J. & Maraschi, L., *Hard TeV spectra of blazars and the constraints to the IR intergalactic background*, MNRAS **368**, L52-L56 (2006). [arXiv:astro-ph/0603030]

Kato, T. *et al.*, *Constraining the Emission Properties of TeV Blazar H1426+428 by the Synchrotron-Self-Compton Model*, ApJ **638** 1, 653-658 (2006). [arXiv:astro-ph/0510730]

Kashlinsky, A., *Cosmic Infrared Background and Early Galaxy Evolution*, Phys. Rep. **409**, 361-438 (2005). [arXiv:astro-ph/0412235]

Kembhavi, A. & Narlikar, J., *Quasars and Active Galactic Nuclei*, Cambridge University Press (1999).

Kerr, R. P., *Gravitational field of a spinning mass as an example of algebraically special metrics*, Phys. Rev. Lett. **11**, 237 (1963).

Kildea, J. *et al.*, *The Whipple Observatory 10 m gamma-ray telescope, 1997-2006*, Astropart. Phys. **28** Iss. 2, 182-195 (2007).

King, A., *Gamma-ray Burst Models*, Proc. Royal Soc. Disc. meet. on GRBs (2006). [arXiv:astro-ph/0609811]

Klebesadel, R. W., Strong, I. B. & Olson, R. A., *Observations of Gamma-Ray Bursts of Cosmic Origin*, ApJL **182**, L85 (1973).

Kneiske, T. M., Mannheim, K. & Hartmann, D. H., *Implications of Cosmological Gamma-Ray Absorption: 1. Evolution of the Metagalactic Radiation Field*, A&A **386**, 1 (2002). [arXiv:astro-ph/0202104]

Kneiske, T. M., Bretz, T., Mannheim, K. & Hartmann, D. H., *Implications of Cosmological Gamma-Ray Absorption: 2. Modification of Gamma-Ray Spectra*, A&A **413**, 807-815 (2004). [arXiv:astro-ph/0309141]

Kneiske, T. M. & Mannheim, K., *BL Lac Contribution to the Extragalactic Gamma-Ray Background*, Proc. 29th ICRC (Pune, India), **4**, 1 (2005). [arXiv:astro-ph/0411146]

Kneiske, T. & Mannheim, K., *BL Lacertae contribution to the extragalactic gamma-ray background*, A&A **479** (1), 41-47 (2008). [arXiv:0705.3778]

Konopelko, A., Mastichiadis, A. *et al.*, *Modeling the TeV Gamma-Ray Spectra of Two Low-Redshift Active Galactic Nuclei: Markarian 501 and Markarian 421*, ApJ **597**, 851-859 (2003).

Konopelko, A. *et al.*, *Modeling the IR de-absorbed γ-ray spectra of TeV BL Lacs*, Proc. 28th ICRC, OG 2.3, 2611-2614 (2003).

Konopelko, A., Mastichiadis, A. & Stecker, F. W., *TeV Gamma-Ray Spectra Unfolded for IR Absorption for a Sample of Low and High Red-Shifted AGN*, Proc. 29th ICRC, Pune, India (2005). [arXiv:astro-ph/0507479]

Konopelko, A., *Stereo Array of 30m Imaging Atmospheric Cherenkov Telescopes: A Next-Generation Detector for Ground-Based High Energy Gamma-ray Astronomy*, Astropart. Phys. **24** (3), 191-207 (2005). [arXiv:astro-ph/0506465]

Kormendy, J. & Gebhardt, K., *Supermassive Black Holes in Nuclei of Galaxies*, Proc. 20th Texas Symp. on Rel. Astroph. **586**, 363 (2001). [arXiv:astro-ph/0105230]

Kotilainen, R. *et al.*, *The optical-near-infrared colour of the host galaxies of BL Lacertae objects*, A&A **424**, 107-118 (2004). [arXiv:astro-ph/0405621]

Koul, R. *et al.*, *The Himalayan Gamma-Ray Observatory at Hanle*, Proc. 29th ICRC (Pune, India), **5**, 243-246 (2005).

Kraushaar, W. L. & Clark, G. W., *Search for primary cosmic gamma rays with the satellite Explorer XI*, PRL **8(3)**, 106-109 (1962).

Krawczynski, H. *et al.*, *Multiwavelength Observations of Strong Flares From the TeV-Blazar 1ES 1959+650*, ApJ **601**, 151-164 (2004). [arXiv:astro-ph/0310158v1]

Kubo, H. *et al.*, *Status of the CANGAROO-III Project*, New Astron. Rev. **48**, 323-329 (2004).

Landau, L. D. & Lifshitz, E. M., *The classical theory of fields*, Course of theoretical physics, Pergamon Press, 3rd rev. ed. (1971).

Laor, A., *Astrophysical evidence for massive black holes*, Phys. Rep. **311**, 451-462 (1999).

LeBohec, S. & Holder, J., *The Cosmic Ray Background as a Tool for Relative Calibration of Atmospheric Cherenkov Telescopes*, Astropart. Phys. **19** (2), 221-233 (2002). [arXiv:astro-ph/0208396]

Levine, A. M. *et al.*, *First Results from the All-Sky Monitor on the Rossi X-Ray Timing Explorer*, ApJ **469**, L33 (1996). [arXiv:astro-ph/9608109]

Levinson, A. & Blandford, R. D., *Pair Cascades in Extragalactic Jets. II. The Beamed X-Ray Spectrum*, ApJ **449**, 86-92 (1995).

Levinson, A., *Pair Cascades in Extragalactic Jets. III. Synchrotron Emission*, ApJ **459**, 520-528 (1996).

Li, L., *Screw Instability and Blandford-Znajek Mechanism*, ApJ **531**, L111-121 (2000). [arXiv:astro-ph/0001420]

Li, T. P. & Ma, Y. Q., *Analysis methods for results in gamma-ray astronomy*, ApJ **272**, 317-324 (1983).

Lindner, T., *Padding*, STACEE Internal Collaboration Note 05-06 (2005).

Lindner, T., *Improvements and Calibration of Grid Alignment Scheme*, STACEE Internal Collaboration Note 06-01 (2006).

Lindner, T., *STACEE Observations of BL Lac Objects*, Ph.D. Thesis, McGill University (2006).

Lindner, T. *et al.*, *Very high energy observations of the BL Lac objects 3C66A and OJ287*, Astropart. Phys. **28**, 338-347 (2007). [arXiv:0707.2815]

Longair, M. S., *High energy astrophysics, Vol. 1: Particles, photons and their detection - 2nd ed.*, Cambridge University Press (1992).

Madau, P. & Pozzetti, L., *Deep galaxy counts, extragalactic background light and the stellar baryon budget*, MNRAS **312** (2), L9-L15 (1999). [arXiv:astro-ph/9907315]

Maiolino, R. *et al.*, *Discovery of a Nuclear Gas Bar Feeding the Active Nucleus in Circinus*, ApJ **531** (1), 219-231 (2000). [arXiv:astro-ph/9910160]

Malkan, M. A. & Stecker, F. W., *An Empirically Based Calculation of the Extragalactic Infrared Background*, ApJ **496**, 13 (1998). [arXiv:astro-ph/9710072]

Malkan, M. A. & Stecker, F. W., *An empirically based model for predicting infrared luminosity functions, infrared galaxy counts, and the diffuse infrared background*, ApJ **555**, 641-649 (2001). [arXiv:astro-ph/0009500]

Mannheim, K., *The proton blazar*, A&A **269**, 67-76 (1993).

Mannheim, K., Westerhoff, S., Meyer, H. & Fink, H.-H., *Beacons at the Gamma Ray Horizon*, A&A **315**, 77-85 (1996). [arXiv:astro-ph/9695107]

Manseri, H., *Astronomie gamma au-dessus de 30 GeV. Une nouvelle méthode d'identification des rayons gamma cosmiques à partir du sol avec le détecteur CELESTE.* Ph.D. thesis, Ecole Polytechnique, Paris (2004).

Mantsch, P., *The Pierre Auger Observatory and first results*, Proc. 29th ICRC, Pune, India (2005). [arXiv:astro-ph/0604114]

Mapelli, M., Salvaterra, R. & Ferrara, A., *Extragalactic Background Light: new constraints from the study of the photon-photon absorption on blazar spectra*, Proc. of Science, BDMH2004, 012 (2004). [arXiv:astro-ph/0411134]

Maraschi, L., Ghisellini, G. & Celotti, A., *A jet model for the gamma-ray emitting blazar 3C 279*, ApJL **397** (1), L5-L9 (1992).

Martin, J.-P. & Ragan, K., *A programmable nanosecond digital delay and trigger system*, Proc. IEEE Nuclear Science Symposium (Lyon, France), **12**, 141 (2000).

Mastichiadis, A. & Kirk, J. G., *Variability in the synchrotron self-Compton model of blazar emission*, A&A **320**, 19-25 (1997). [arXiv:astro-ph/9610058]

Mazin, D., *Einschränkungen der Energiedichte des extragalaktischen Hintergrundlichtes mit Hilfe der gemessenen TeV-γ-Strahlung*, Ph.D. thesis, Institut für Experimentalphysik, Universität Hamburg (2003).

Mazin, D., *Observations of Extragalactic Sources with the MAGIC Telescope*, Astrophys. Space Sci. (2006). [arXiv:astro-ph/0609152]

Mazin, D. & Goebel, F., *Break in the VHE spectrum of PG 1553+113: new Upper Limit on its redshift?*, ApJ **655**, L13-L16 (2006). [arXiv:astro-ph/0611817]

Mazin, D. & Raue, M., *New limits on the density of the extragalactic background light in the optical to the far-infrared from the spectra of all known TeV blazars*, A&A **471** (2), August IV 2007, 439-452 (2007). [arXiv:astro-ph/0701694]

Misner, C. W., Thorne, K. S. & Wheeler, J. A., *Gravitation*, San Francisco: Freeman (1973).

Morrison, P., *On gamma-ray astronomy*, Nuovo Cimento **7(6)**, 858-865 (1958).

Mücke, A. & Protheroe, R. J., *A proton synchrotron blazar model for flaring in Markarian 501*, Astropart. Phys. **15** (1), 121-136 (2001). [arXiv:astro-ph/0004052v3]

Mücke, A., Protheroe, R. J. *et al.*, *BL Lac Objects in the Synchrotron Proton Blazar Model*, Astropart. Phys. **18**, 593-613 (2002). [arXiv:astro-ph/0206164]

Mueller, C., *Simulations of STACEE's Canting Behavior*, STACEE Internal Collaboration Note 04-04 (2004).

Mueller, C., *CORSIKA for STACEE*, STACEE Internal Collaboration Note 05-02 (2005).

Mueller, C., *Trigger and coincidence gate width studies*, STACEE Internal Collaboration Note 05-05 (2005).

Narayan, R. & Yi, I., *Advection-dominated accretion: A self-similar solution*, ApJ **428**, L13-L16 (1994). [arXiv:astro-ph/9403052]

Nayakshin, S., *The physics of accretion disks with magnetic flares*, Ph.D. thesis, University of Arizona (1998). [arXiv:astro-ph/9811061]

Nelson, W. R., Hirayama, H. & Rogers, D. W. O., Rep. SLAC 265, Stanford Linear Accelerator Center (1985).

Neronov, A. & Semikoz, D. V., *Measurement of extragalactic magnetic fields by TeV gamma ray telescopes*, JETP Lett. **85**, 579 (2007). [arXiv:astro-ph/0604607]

Neshpor, Iu.I. *et al.*, *On the periodical very high energy gamma-ray emission from Cyg X-3*, Astrophys. Space Sci. **61** (2), 349-355 (1979).

Ning, X., Winston, R. & O'Gallagher, J., *Dielectric totally internally reflecting concentrators*, Applied Optics **26**, 300-305 (1987).

Oda, H., Machida, M., Nakamura, K. E. & Matsumoto, R., *Steady Models of Optically Thin, Magnetically Supported Black Hole Accretion Disks*, Publ. of the Astron. Soc. of Japan **59** (2), 457-465 (2007). [arXiv:astro-ph/0701658]

Ong, R. A., *Very high-energy gamma-ray astronomy*, Phys. Rep. **305**, 93-202 (1998).

Ong, R. A., *The Status of VHE Gamma-Ray Astronomy*, arXiv:astro-ph/0605191 (2006).

Ormes, J., *GLAST: Progress and Plans as of March 2003*, Proc. APS HEAD 2003, Mt. Tremblant, Canada (2003).

Oser, S. M., *Analysis of Heliostat Sunspot Images and Comparison to Simulations*, STACEE Internal Collaboration Note 99-08 (1999).

Oser, S. M., *High Energy Gamma-Ray Observations of the Crab Nebula and Pulsar with the Solar Tower Atmospheric Cherenkov Effect Experiment*, Ph.D. thesis, University of Chicago (2000).

Ostapchenko, S., *QGSJET-II: results for extensive air showers*, Proc. XIII ISVHECRI, Pylos, Greece (2004). [arXiv:astro-ph/0412591]

Padovani, P. & Giommi, P., *The connection between x-ray- and radio-selected BL Lacertae objects*, ApJ **444**, 567-581 (1995).

Paré, E. *et al.*, *CELESTE: An atmospheric Cherenkov telescope for high energy gamma astrophysics*, NIM A **490**, 71-89 (2002).

Perkins, D., *Particle Astrophysics*, Oxford University Press (2003).

Perlman, E. S., Biretta, J. A., Zhou, F., Sparks, W. B. & Macchetto, D., *Optical and Radio Polarimetry of the M87 Jet at 0.2" Resolution*, Astron. J. **117**, 2185-2198 (1999). [arXiv:astro-ph/9901176v1]

Petry, D. *et al.*, *The TeV Spectrum of H1426+428*, ApJ **580** 1, 104-109 (2002). [arXiv:astro-ph/0207506]

Photonis Corp., *Photomultiplier Tubes: 2004 Catalogue*, Photonis, Brive, France (2004).

Piner, B. G., Pant, N. & Edwards, P. G., *The Parsec-Scale Jets of the TeV Blazars H 1426+428, 1ES 1959+650, and PKS 2155-304: 2001-2004*, ApJ , *in press* (2008). [arXiv:0801.2749v1]

Piron, F. *et al.*, *Temporal and spectral gamma-ray properties of Mkn 421 above 250 GeV from CAT observations between 1996 and 2000*, A&A **374**, 895-906 (2001). [arXiv:astro-ph/0106196]

Pittori, C. & Tavani, M., *Scientific Goals and Instrument Performance of the Gamma-Ray Imaging Detector AGILE*, Nucl. Phys. B (Proc. Suppl.) **134**, 72-74 (2004).

Plaga, R., *Detecting intergalactic magnetic fields using time delays in pulses of big gamma-rays*, Nature **374**, 430-432 (1995).

Prahl, J., *Suche nach kosmischen γ-Punktquellen oberhalb von 20 TeV mit den HEGRA-Detektorfeldern*, Ph.D. thesis, II. Institut für Experimentalphysik, Universität Hamburg (1999).

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P., *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition, Cambridge University Press, Cambridge (1992).

Primack, J. R., Bullock, J. S., Somerville, R. S. & MacMinn, D., *Probing galaxy formation with TeV gamma ray absorption*, Astropart. Phys. **11**, 93-102 (1999).

Primack, J. R., Somerville, R. S. *et al.*, *Probing Galaxy Formation with High Energy Gamma-Rays*, AIP Conf. Proc. **558**, 463-478 (2001). [arXiv:astro-ph/0011475]

Primack, J. R., *Probing Galaxy Formation with High Energy Gamma-Rays*, Frascati Physics Series **XXIV**, 559-580 (2002). [arXiv:astro-ph/0201119]

Primack, J. R., *et al.*, *Observational Gamma-ray Cosmology*, AIP Conf. Proc. **745**, 23 (2005). [arXiv:astro-ph/0502177]

Ragan, K., *The Montreal digital delay system Specifications and Design*, STACEE Internal Collaboration Note 03-01 (2003).

Ramana Murthy, P. V. & Wolfendale, A. W., *Gamma-Ray Astronomy*, Cambridge University Press, Cambridge (1993).

Rees, M. J., Begelman, M. C., Blandford, R. D. & Phinney, E. S., *Ion-supported tori and the origin of radio jets*, Nature **295**, 17-21 (1982).

Remillard, R. A. *et al.*, *Two X-Ray Selected BL Lacertae Objects Observed with the HEAO1 Scanning Modulation Collimator*, ApJ **345**, 140-147 (1989).

Sambruna, R. M. *et al.*, *An X-Ray Absorption Feature in the BL Lacertae Object H1426+428*, ApJ **483**, 774 (1997).

Sambruna, R. M. & Mushotzky, R. F., *X-Ray Absorption in the BL Lacertae Object PKS 0548-322*, ApJ **502**, 630 (1998).

Sambruna, R. M., Donato, D., Tavecchio, F. *et al.*, *Deep Chandra and multicolor HST observations of the jets of 3C 371 and PKS 2201+044*, ApJ **670** (1), 74-91 (2007). [arXiv:0707.3980]

Saz Parkinson, P. M. *et al.*, *Detection of TeV Gamma-Rays from extended sources with Milagro*, Proc. 22nd Texas Symp. on Rel. Astroph. (2004). [arXiv:astro-ph/0503244]

Scalzo, R. A., *Simulations of Heliostat Defocusing*, STACEE Internal Collaboration Note 00-08 (2000).

Scalzo, R. A., *The STACEE Atmospheric Monitor: Status and Plans for Expansion*, STACEE Internal Collaboration Note 00-13 (2000).

Scalzo, R. A., *A New CCD Autoguiding System for the STACEE Atmospheric Monitor*, STACEE Internal Collaboration Note 02-04 (2002).

Scalzo, R. A., *Poisson Template Methods with Undercanting: Fitting Shower Energies and Core Locations*, STACEE Internal Collaboration Note 02-05 (2002).

Scalzo, R. A., *Advanced Reconstruction I: Hadron Rejection Using Monte Carlo Templates*, STACEE Internal Collaboration Note 02-15 (2002).

Scalzo, R. A. *et al.*, *Optimized pointing strategies for solar tower ACTs*, Proc. 28th ICRC (Tsukuba, Japan), OG 2.5, p. 2799-2802 (2003).

Scalzo, R. A., *Observations of the EGRET blazar W Comae with the Solar Tower Atmospheric Cherenkov Effect Experiment*, Ph.D thesis, University of Chicago (2004).

Scarpa, R., Urry, C. M. *et al.*, *The Hubble space telescope survey of BL Lacertae objects. I. Surface brightness profiles, magnitudes, and radii of host galaxies*, ApJ **532** (2), 740-815 (2000). [arXiv:astro-ph/9911147]

Schlenker, S., Lohse, T., Borgmeier, C. & Stegmann, C., *Simulation of Night Sky Background Trigger Rate for the H.E.S.S. Camera*, HESS internal note, Humboldt Universität Berlin (2001).

Schlickeiser, R., *Cosmic Ray Astrophysics*, Springer (2003).

Schroedter, M., *Upper Limits on the Extragalactic Background Light from the Very High Energy Gamma-Ray Spectra of Blazars*, ApJ **628** (2), 617-628 (2005). [arXiv:astro-ph/0504397]

Schuette, D. R., *Development of an Electronics System for Precision Atmospheric Cherenkov Effect Reconstruction*, Undergraduate senior thesis, University of Chicago (2000).

Shakura, N. I. & Sunyaev, R. A., *Black holes in binary systems. Observational appearance*, A&A **24**, 337-355 (1973).

Sikora, M., Begelman, M. C., Madejski, G. M. & Lasota, J.-P., *Are Quasar Jets Dominated by Poynting Flux?*, ApJ **625**, 72-77 (2005).

Smith, A. J., *The MILAGRO Gamma-Ray Observatory*, Proc. 29th ICRC (Pune, India), **10**, 227-242 (2005).

Smith, D. A. *et al.*, *Mrk 421, Mrk 501, and 1ES 1426+428 at 100 GeV with the CELESTE Cherenkov telescope*, A&A **459** (2), 453-464 (2006). [arXiv:astro-ph/0608247]

Smith, P. S. *et al.*, *UBVRI photometry of stars in the fields of X-ray selected BL Lacertae objects*, ApJS **77**, 67-73 (1991).

Spada, M., Ghisellini, G., Lazzati, D. & Celotti, A., *Internal shocks in the jets of radio-loud quasars*, MNRAS **325**, 1559-1570 (2001). [arXiv:astro-ph/0103424]

Stecker, F. W., De Jager, O. C. & Salamon, M. H., *TeV gamma rays from 3C 279 - A possible probe of origin and intergalactic infrared radiation fields*, ApJL **390**, L49-L52 (1992).

Stecker, F. W., Malkan, M. A. & Scully, S. T., *Intergalactic Photon Spectra from the Far-IR to the UV Lyman Limit for $0 < z < 6$ and the Optical Depth of the Universe to High-Energy Gamma Rays*, ApJ **648** (2), 774-783 (2006). [arXiv:astro-ph/0510449]

Stecker, F. W. & Scully, S. T., *A Simple Analytic Treatment of the Intergalactic Absorption Effect in Blazar Gamma-Ray Spectra*, ApJL **652**, L9 (2006). [arXiv:astro-ph/0608110]

Stecker, F. W., Baring, M. G. & Summerlin, E. J., *Blazar gamma-rays, shock acceleration, and the Extragalactic Background Light*, ApJL **667**, L29 (2007). [arXiv:0707.4676v2]

Stecker, F. W. & Scully, S. T., *The Spectrum of 1ES0229 + 200 and the Cosmic Infrared Background*, subm. to A&A (2007). [arXiv:0710.2252]

Stickel, M., Fried, J. W. & Kühr, H., *The complete sample of 1 Jy BL Lac objects. II - Observational data*, A&AS **98**, 393-442 (1993).

Takeda, M. *et al.*, *Energy determination in the Akeno Giant Air Shower Array Experiment*, Astropart. Phys. **19**, 447-462 (2003).

Tavecchio, F., Maraschi, L. & Ghisellini, G., *Constraints on the Physical Parameters of TeV Blazars*, ApJ **509**, 608-619 (1998).

Tavecchio, F., Maraschi, L. & Ghisellini, G., *Theoretical Implications from the Spectral Evolution of Markarian 501 Observed with BeppoSAX*, ApJ **554**, 725-733 (2001).

Teshima, M. *et al.*, *Discovery of Very High Energy Gamma-Rays from the Distant Flat Spectrum Radio Quasar 3C 279 with the MAGIC Telescope*, Proc. 30th ICRC, Mérida, Mexico (2007). [arXiv:0709.1475]

Thompson, D. J. *et al.*, *Calibration of the Energetic Gamma-Ray Experiment Telescope (EGRET) for the Compton Gamma-Ray Observatory*, ApJS **86**, 629-656 (1993).

Thomson, G. B., *Observation of the GZK Cutoff by the HiRes Experiment*, Proc. Quarks'06 (2006). [arXiv:astro-ph/0609403]

Tibet AS-Gamma Collaboration, *A Northern Sky Survey for Steady Tera-Electron Volt Gamma-Ray Point Sources Using the Tibet Air Shower Array*, ApJ **633**, 1005-1012 (2005).

Tinyakov, P. G. & Tkachev, I. I., *Correlation function of ultra-high energy cosmic rays favors point sources*, JETP Lett. **74**, 445 (2001). [arXiv:astro-ph/0102101]

Tramacere, A. *et al.*, *SWIFT observations of TeV BL Lacertae objects*, A&A **467**, 501-508 (2007). [arXiv:astro-ph/0611276]

Urry, C. M. & Padovani, P., *Unified Schemes for Radio-Loud Active Galactic Nuclei*, PASP **107** 715, 803-845 (1995). [arXiv:astro-ph/9506063]

Urry, C. M. *et al.*, *The HST Survey of BL Lacertae Objects. II. Host Galaxies*, ApJ **532** (2), 816-829 (2000). [arXiv:astro-ph/9911109]

Vincent, P. *et al.*, Proc. 29th ICRC (Pune, India), **5**, 163 (2005).

Wagner, R., *Extragalactic Jets: Theory and Observation from Radio to Gamma Ray*, ASP Conf. Ser. (in press, 2007). [arXiv:0706.4439]

Weekes, T. C. & Turver, K. E., Proc. 12th ESLAB Symposium (Frascati, Italy), 279 (1977).

Weekes, T. C., *Very high energy gamma-ray astronomy*, Phys. Rep. **160**, 1-2 (1988).

Weekes, T. C. *et al.*, *Observation of TeV gamma rays from the Crab Nebula using the atmospheric Cerenkov imaging technique*, ApJ **342**, 379-395 (1989).

Westerhoff, S. *et al.*, *HiRes: Searching for the origins of ultra high energy cosmic rays*, AIP Conf. Proc. **698**, 370-373 (2004).

Wilson, A. S. & Colbert, E. J. M., *The difference between radio-loud and radio-quiet active galaxies*, ApJ **438** (1), 62-71 (1995). [arXiv:astro-ph/9408005]

Wischnewski, R., *The BAIKAL Neutrino Experiment: From NT200 to NT200+*, Proc. VLVNT2, Catania, Italy (2005). [arXiv:astro-ph/0609743]

Wolter, A. *et al.*, *The hard synchrotron X-ray spectrum of the TeV BL Lac 1ES 1426+428*, Proc. Extragalactic Jets (Girdwood, Alaska), May 21-24 (2007). [arXiv:0707.2735]

Wood, K. S. *et al.*, *The HEAO A-1 X-ray source catalog*, ApJS **56**, 507-649 (1984).

Wu, X.-B. *et al.*, *Supermassive black hole masses of AGNs with elliptical hosts*, A&A **389**, 742-751 (2002).

Zatsepin, G. T. & Kuzmin, V. A., *Upper Limit of the Spectrum of Cosmic Rays*, JETP Lett. **4**, 78 (1966); Zh. Exp. Theor. Phys. Pis ma Redaktsiiu **4**, 114 (1966).

Zweerink, J., private communication (2004).

Web Sites

CGRO web site (2006),
`http://cossc.gsfc.nasa.gov/`

Chandra X-ray Center (CXC) at Harvard Univ. (2008),
`http://cxc.harvard.edu/newsletters/news_13/jets.html`

CORSIKA web site (2008),
`http://www-ik.fzk.de/corsika/`

FLUKA web site (2008),
`http://www.fluka.org`

GLAST web site (2006),
`http://glast.stanford.edu/`

HEGRA web site (2006),
`http://www.mpi-hd.mpg.de/hfm/CT/CT.html`

NRC-CNRC (EGS4) web page (2008),
`http://www.irs.inms.nrc.ca/EGS4/get_egs4.html`

Photonis web site (2005),
`http://www.photonis.com`

ROSAT web site (Max-Planck-Institut für extraterrestrische Physik; 2006),
`http://wave.xray.mpe.mpg.de/rosat`

RXTE-ASM web site (2007),
`http://xte.mit.edu/ASM_lc.html`

SOHO web site (2006),
`http://sohowww.nascom.nasa.gov/`

Wagner, R., Web page on ground-based gamma-ray astronomy (2007),
`http://www.mppmu.mpg.de/~rwagner/sources/`