

**Application of Artificial Neural Network methods for improving  
the sensitivity of TACTIC Cherenkov telescope**

*By*

**VIR KRISHEN DHAR**

**PHYS01200704010**

**Bhabha Atomic Research Centre, Mumbai**

*A thesis submitted to the*

*Board of Studies in Physical Sciences*

*In partial fulfillment of requirements*

*For the Degree of*

**DOCTOR OF PHILOSOPHY**

*of*

**HOMI BHABHA NATIONAL INSTITUTE**



**2013**

# Homi Bhabha National Institute

## Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Vir Krishen Dhar entitled "Application of Artificial Neural Network methods for improving the sensitivity of TACTIC Cherenkov telescope", and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.



Date: 2/12/2013

Chairman - <Dr B S Acharya>



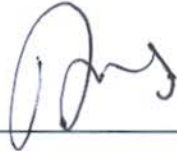
Date:

Guide / Convener - <Dr A K Tickoo> 2/12/13



2.12.13. Date:

Member 1 - <Prof. N K Gupta>



2.12.13. Date:

Member 2 - <Prof A K Mohanty>



Date:

Member 3 - <Mr R Koul>

2.12.13.

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

  
(A K Tickoo)

Date: 4.12.2013.

Place: Mumbai

## **STATEMENT BY AUTHOR**

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Vir Krishen Dhar

## **DECLARATION**

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Vir Krishen Dhar

## List of papers in refereed journals:(After Registration)

---

- [1]. Artificial Neural Network based gamma-hadron segregation methodology for TACTIC telescope.  
**V.K.Dhar**, A.K.Tickoo, M.K.Koul, R.Koul, B.P.Dubey, R.C.Rannot, K.K.Yadav, P.Chandra, M.Kothari, K.Chanchalani, K.Venugopal.  
*Nucl. Instrum. and Meth. A.*, 708 (2013) 56.
- [2]. TeV gamma-ray observations of Markarian 421 using TACTIC during 2009-2010.  
P. Chandra, R. C. Rannot, K.K.Yadav, A.K.Tickoo, K.K.Singh, K.Chanchalani, M.Kothari, N.K.Agarwal, A.Goyal, H.C.Goyal, S.Kotwal, N.Kumar, P.Marandi, K.Venugopal, C.K.Bhat, N.Bhatt, S. Bhattacharyya, C.Borwankar, N.Chouhan, **V.K.Dhar**, S.R.Kaul, S.K.Koul, M.K.Koul, R.Koul, A.K.Mitra, S.Sahaynathan and M.Sharma.  
*J.Phys. G: Nucl. Part. Phys.*, 39 (2012) 045201.
- [3]. Simulation studies for optimizing the trigger generation criteria for the TACTIC telescope  
M.K.Koul, A.K.Tickoo, **V.K.Dhar**, K.Venugopal, K.Chanchalani, R.C.Rannot, K.K.Yadav, P.Chandra, M.Kothari, R.Koul  
*Nucl. Instrum. and Meth. A.*, 646 (2011) 204.
- [4]. TeV observations of Mrk 421 with the TACTIC  $\gamma$ -ray telescope during 2006-2008.  
P. Chandra, K.K.Yadav, R. C. Rannot, K.K.Singh, A.K.Tickoo, M. Sharma, K.Venugopal, C.K.Bhat, N. Bhatt, S. Bhattacharyya, K.Chanchalani, **V.K.Dhar**, S.V.Godambe, H.C.Goyal, M.Kothari, S.Kotwal, M.K.Koul, R. Koul and S.Sahaynathan.  
*J.Phys. G: Nucl. Part. Phys.*, 37 (2010) 125201.
- [5]. Comparative performance of some popular ANN algorithms on benchmark and function approximation problems  
**V.K.Dhar**, A.K.Tickoo, R.Koul, B.P.Dubey .  
*Pramana- Journal of Phys.*, 74 (2010) 307.
- [6]. Artificial Neural Network-based error compensation procedure for low-cost encoders.  
**V.K.Dhar**, A.K.Tickoo, S.K.Kaul, R.Koul, B.P.Dubey .  
*Meas. Sci. Technol.*, 21 (2010) 015112.
- [7]. ANN-based energy reconstruction procedure for TACTIC gamma-ray telescope and its comparison with other conventional methods  
**V.K.Dhar**, A.K.Tickoo, M.K.Koul, R.C.Rannot, K.K.Yadav, P.Chandra, B.P.Dubey, R.Koul  
*Nucl. Instrum. and Meth. A.*, 606 (2009) 795.
- [8]. Search for TeV  $\gamma$ -ray from H1426+428 during 2004-2007 with TACTIC telescope.

K.K.Yadav, R. C. Rannot, P. Chandra, A.K.Tickoo, S.Thoudam, K.Venugopal, N. Bhatt, S. Bhattacharyya, K.Chanchalani, **V.K.Dhar**, S.V.Godambe, H.C.Goyal, M.Kothari, S.Kotwal, M.K.Koul, R. Koul, S.Sahaynathan, M. Sharma.

*J.Phys. G: Nucl. Part. Phys.*, 36 (2009) 085201.

[9]. Very High Energy  $\gamma$ -ray observations of Mrk 501 using TACTIC imaging  $\gamma$ -ray telescope during 2005-06.

S. V. Godambe, R. C. Rannot, P. Chandra, K.K.Yadav, A.K.Tickoo, K.Venugopal, N. Bhatt, S. Bhattacharyya, K.Chanchalani, **V.K.Dhar**, H.C.Goyal, R.K.Kaul, M.Kothari, S.Kotwal, M.K.Koul, R. Koul, S.Sahaynathan, M. Sharma, S.Thoudam.

*J.Phys. G: Nucl. Part. Phys.*, 35 (2008) 065202.

[10]. The TACTIC atmospheric Cherenkov Imaging telescope

R.Koul, A.K.Tickoo, S.K.Kaul, S.R.Kaul, N.Kumar, K.K.Yadav, N.Bhatt, K.Venugopal, H.C.Goyal, M.Kothari, P.Chandra, R.C.Rannot, **V.K.Dhar**, M.K.Koul, R.K.Kaul, S.Kotwal, K.Chanchalani, S.Thoudam, N.Chouhan, M.Sharma, S.Bhattacharyya, S.Sahayanathan.

*Nucl. Instrum. and Meth. A.*, 578 (2007) 548.

[11]. Very High Energy  $\gamma$ -ray and Near Infrared observations of 1ES2344+514 during 2004-05.

S.V.Godambe, R.C.Rannot, K.S.Baliyan, A.K.Tickoo, S.Thoudam, **V.K.Dhar**, P.Chandra, K.K.Yadav, K.Venugopal, N.Bhatt, S.Bhattacharyya, K.Chanchalani, S.Ganesh, H.C.Goyal, U.C.Joshi, R.K.Kaul, M.Kothari, S.Kotwal, M.K.Koul, R.Koul, S.Sahayanathan, C.Shah and M.Sharma.

*J.Phys. G: Nucl. Part. Phys.*, 34 (2007) 1683.

[12]. Observations of TeV  $\gamma$ -rays from Mrk 421 during December 2005 to April 2006 with the TACTIC telescope.

K.K.Yadav, P.Chandra, A.K.Tickoo, R.C.Rannot, S.Godambe, M.K.Koul, **V.K.Dhar**, S.Thoudam, N.Bhatt, S.Bhattacharyya, K.Chanchalani, H.C.Goyal, R.K.Kaul, M.Kothari, S.Kotwal, R.Koul, S.Sahayanathan, M.Sharma and K.Venugopal.

*Astropart. Phys.*, 27 (2007) 447.

## List of papers in Conference proceedings:(After Registration)

[1]. Sensitivity improvement of the TACTIC  $\gamma$ -ray telescope.

A. K. Tickoo, S.R.Kaul K.Venugopal, N.K.Agarwal, H.C.Goyal, P.Chandra, N.Chouhan, **V.K.Dhar**, K.K.Yadav, R.C.Rannot, R.Koul, K.K.Singh, N.Kumar, P.Marande, K.Chanchlani, K.K.Gour, M.Kothari, A.Goyal, A.K.Mitra, B.S.Sahayanathan, M.Sharma, C.K.Bhat, N.Bhatt, S.Bhattacharyya, C.Borwankar, M.K.Koul.

*Presented at “ TeV Particle Astrophysics 2012 ”; Mumbai, India,(2012).*

[2]. Recent TeV observations of Mrk 501, Mrk 421 and M87 with TACTIC.

R.C.Rannot, P.Chandra, K.K.Yadav, H.Bhatt, A. K. Tickoo, K.K.Singh, A.Goyal, H.C.Goyal, N.Kumar, P.Marandi, N.K.Agarwal, K.Chanchalani, K.K.Gour, M.Kothari, N.Chouhan, **V.K.Dhar**, S.R.Kaul, M.K.Koul, R.Koul, A.K.Mitra, S.Sahayanathan, M.Sharma, K.Venugopal, C.K.Bhat, N.Bhatt, S.Bhattacharyya, C.Borwankar.

*Presented at “ TeV Particle Astrophysics 2012 ”; Mumbai, India,(2012).*

[3]. Simulating studies for optimizing the trigger field of view of the TACTIC telescope.

M.K.Koul, A.K.Tickoo, **V.K.Dhar**, K.Venugopal, K.Chanchalani, R.C.Rannot, K.K.Yadav, P.Chandra, M.Kothari, R.Koul

*Presented at “ 29<sup>th</sup> ASI meeting”; Raipur, India, (2011)*

[4]. Mrk-421 in a high TeV emission state during 2010 : TACTIC observations:

P.Chandra, R.C.Rannot, K.K.Yadav, A.K.Tickoo, K.Chanchalani, M.Kothari, K.K.Singh, N.K.Agarwal, A.Goyal, S.Kotwal, N.Kumar, P.Marandi, K.Venugopal, C.K.Bhat, N.Bhatt, S.Bhattacharyya, N.Chouhan, **V.K.Dhar**, S.R.Kaul, S.K.Koul, M.K.Koul, R.Koul, A.K.Mitra, S.Sahayanathan and M.Sharma.

*Presented at “ 29<sup>th</sup> ASI meeting”; Raipur, India, (2011)*

[5]. VHE gamma-ray observations of Mrk-501 and 1ES2344+514 with TACTIC during 2009-2010.

M.Kothari, P.Chandra, R.C.Rannot, K.K.Yadav, A.K.Tickoo, K.Chanchalani, K.K.Singh, K.Venugopal, N.K.Agarwal, A.Goyal, H.C.Goyal, S.Kotwal, N.Kumar, P.Marandi, M.Sharma, C.K.Bhat, N.Bhatt, S.Bhattacharyya, N.Chouhan, **V.K.Dhar**, S.R.Kaul, S.K.Koul, M.K.Koul, R.Koul, A.K.Mitra and S.Sahayanathan.

*Presented at “ 29<sup>th</sup> ASI meeting”; Raipur, India, (2011)*

[6]. Feasibility of operating TACTIC telescope during partial moonlit conditions .

K.K.Singh, A.K.Tickoo, K.K.Yadav, R.C.Rannot, **V.K.Dhar** and R.Koul.

*Presented at “ 16<sup>th</sup> National Space Science Symposium ( NSSS-2010)”, Rajkot, India, (2010)*

[7]. Recent TeV observations of Mrk-421 with the TACTIC gamma-ray telescope.

R.C.Rannot, P.Chandra, K.K.Yadav, A.K.Tickoo, K.Chanchalani, A.Goyal, H.C.Goyal, M.Kothari, S.Kotwal, N.Kumar, P.Marandi, K.Venugopal, C.K.Bhat, N.Bhatt, S.Bhattacharyya, **V.K.Dhar**,

M.K.Koul, R.Koul, S.Sahayanathan, M.Sharma and K.K.Singh.

*Presented at “ 16<sup>th</sup> National Space Science Symposium ( NSSS-2010)”, Rajkot, India, (2010)*

.....  
[8]. VHE gamma-ray observations of 3C279 during 2008-09 using TACTIC telescope.

P.Chandra, R.C.Rannot, K.K.Yadav, A.K.Tickoo, K.Venugopal, K.Chanchalani, A.Goyal, H.C.Goyal, M.Kothari, S.Kotwal, N.Kumar, P.Marandi, K.K.Singh, C.K.Bhat, N.Bhatt, S.Bhattacharyya, **V.K.Dhar**, M.K.Koul, R.Koul, S.Sahayanathan and M.Sharma.

*Presented at “ 16<sup>th</sup> National Space Science Symposium ( NSSS-2010)”, Rajkot, India, (2010)*

.....  
[9]. TeV gamma-ray observations of the Crab Nebula with TACTIC telescope.

A.K.Tickoo, R.C.Rannot, P.Chandra, K.K.Yadav, **V.K.Dhar**, M.K.Koul, S.Thoudam, M.Sharma, C.K. Bhat, N.Bhatt, S.Bhattacharyya, K.Chanchalani, S.V.Godambe, H.C.Goyal, M.Kothari, S.Kotwal, R.Koul, S.Sahayanathan and K.Venugopal.

*Presented at “ 27<sup>th</sup> ASI meeting”; Bangalore, India, (2009)*

.....  
[10]. TeV gamma-ray observations of Mrk 421 during 2007-08 with TACTIC telescope.

K.K.Yadav, R.C.Rannot, P.Chandra, A.K.Tickoo, **V.K.Dhar**, M.K.Koul, S.Thoudam, S.Sahayanathan, M.Sharma, C.K.Bhat, N.Bhatt, S.Bhattacharyya, K.Chanchalani, S.V.Godambe, H.C.Goyal, M.Kothari, S.Kotwal, R.Koul and K.Venugopal.

*Presented at “ 27<sup>th</sup> ASI meeting”; Bangalore, India, (2009)*

.....  
[11]. The TACTIC atmospheric Cherenkov Imaging telescope

R.Koul, A.K.Tickoo, S.K.Kaul, S.R.Kaul, N.Kumar, K.K.Yadav, N.Bhatt, K.Venugopal, H.C.Goyal, M.Kothari, P.Chandra, R.C.Rannot, **V.K.Dhar**, M.K.Koul, R.K.Kaul, S.Kotwal, K.Chanchalani, S.Thoudam, N.Chouhan, M.Sharma, S.Bhattacharyya, S.Sahayanathan.

*Presented at “ National Symposium on Gamma-ray Astronomy”; Banglore, India, (2007)*

.....  
[12]. Observations of Mrk 421 and Mrk 501 during 2005-2006 with TACTIC telescope.

K.K.Yadav, P.Chandra, A.K.Tickoo, R.C.Rannot, S.V.Godambe, M.K.Koul, **V.K.Dhar**, S.Thoudam, N.Bhatt, S.Bhattacharyya, K.Chanchalani, H.C.Goyal, R.K.Kaul, M.Kothari, S.Kotwal, R.Koul, S.Sahayanathan, M.Sharma and K.Venugopal.

*Presented at “ National Symposium on Gamma-ray Astronomy”; Banglore, India, (2007)*

.....  
[13]. Feasibility of using ANN-based algorithms for improving the sensitivity of TACTIC  $\gamma$ -ray telescope.

**V.K.Dhar**, A.K.Tickoo, M.K.Koul, R.Koul and B.P.Dubey

*Presented at “ 25<sup>th</sup> ASI meeting”; Hyderabad, India, (2007)*

**Selected for the Best Poster award**

.....  
[14]. TACTIC observations of Mrk 421 at TeV energies during December, 2005 to April, 2006.

K.K.Yadav, P.Chandra, A.K.Tickoo, R.C.Rannot, S.V.Godambe, M.K.Koul, **V.K.Dhar**,



S.Thoudam, N.Bhatt, S.Bhattacharyya, K.Chanchalani, H.C.Goyal, R.K.Kaul, M.Kothari, S.Kotwal, R.Koul, S.Sahayanathan, M.Sharma and K.Venugopal.

*Presented at “ 25<sup>th</sup> ASI meeting”; Hyderabad, India, (2007)*

.....  
[15]. TeV observations of Mrk 501 with TACTIC  $\gamma$ -ray telescope during 2005-2006.

S.V.Godambe, P.Chandra, R.C.Rannot, K.K.Yadav, A.K.Tickoo, K.Venugopal, N.Bhatt, S.Bhattacharyya, **V.K.Dhar**, H.C.Goyal, R.K.Kaul, M.Kothari, S.Kotwal, M.K.Koul, R.Koul, S.Sahayanathan, M.Sharma and S.Thoudam.

*Presented at “ 25<sup>th</sup> ASI meeting”; Hyderabad, India, (2007)*

.....

---

## DEDICATIONS

*Memories left behind by Late Dr B B Bhan and Vishal Kaul will always be cherished.*

*I dedicate this thesis to the everlasting memory of my father Late Dr. P. K. Dhar.*

*I am sure you can feel my heart.*

## ACKNOWLEDGEMENTS

This thesis marks the end of a journey in acquiring my Ph.D. It would not have been possible to see the completion of the thesis without the support and encouragement of my friends, colleagues and my worthy institution.

At this moment of accomplishment, first of all I thank my guide and convener doctoral committee, Dr A.K. Tickoo. This work would not have been possible without his guidance, support and encouragement. Under his able guidance I successfully overcame many obstacles and learned a lot from his knowledge, wisdom and expertise. His unflinching conviction will always inspire me and I hope to have the privilege of continuing to work with him for many years to come.

I am extremely indebted to Mr. Ramesh Koul, Head ApSD for allowing me to work in this interesting field. I express my sincere gratitude to him for motivating me to pursue my Ph.D and also for critically examining the thesis related publications. I warmly thank him for his extensive suggestions related to my work.

I gratefully acknowledge the support and encouragement provided by my Ph.D doctoral committee members, Prof B.S. Acharya, Dr N.K. Gupta and Dr A.K. Mohanty. Despite their hectic schedules they have been kind enough to make themselves available for useful discussions. Their suggestions have vastly helped me to improve the quality of this work.

I thank Dr R.C. Rannot, Kuldeep, H.C.Goyal, Pradeep, Anita, Mahendra, Naveen, Prashant Nandan, Mradul, Sunder, Sagar and Dr CK Bhat for discussions on various aspects of the work including observations and data analysis. Help received from Maharaj for providing simulation inputs is duly acknowledged. My sincere thanks are due to Sh. S.R. Kaul, Nilesh, Venugopal, Krishna, Kishore, Chinmay and Sandeep for their contributions at various stages of this work and also for allowing me to vent out the frustration over innumerable cups of tea, especially when things did not happen my way. Thanks for being patient and for tolerating me. Support of Nilay and Subir is acknowledged for their suggestions at various stages of my research.

Thanks are also due to Dr B.P. Dubey, Dr P.V. Varde, Dr Archana Sharma, Dr D. Dutta and other members of BIKAS team, for elaborate technical discussions on various aspects of ANN. Thanks for doing all the explaining.

Dr A.K. Mitra, Dr A.K. Razdan and Dr D.K. Koul are thanked for numerous scientific, political and spiritual discussions which rejuvenated me on various occasions.

Sh Sunil Ganju and Sh Ramesh Raina deserve a special mention for keeping me spirited enough to fight on, during our various encounters. Thanks for all those special moments.

I am indebted to my mother Savitri, from whom I learnt to fight on. Her fighting spirit that has taught me to keep going the when chips are down. Thanks are also due to my sister Smt Veena and my brother Sh B K Dhar for always standing by me wholeheartedly.

Last but not the least, I thank my wife Smriti, and my kids Sarthak and Ojasvi for being there for me and thank them wholeheartedly for their love and support.

Vir Krishen Dhar.



## **Homi Bhabha National Institute**

### **Ph. D. PROGRAMME**

1. **Name of the Student:** V K Dhar
2. **Name of the Constituent Institution:** Bhabha Atomic Research Centre
3. **Enrolment No. :** PHYS01200704010
4. **Title of the Thesis:** Application of ANN methods for improving the sensitivity of TACTIC Cherenkov telescope.
5. **Board of Studies:** Physical Sciences

### **SYNOPSIS**

Very High energy (VHE) gamma-rays in the energy range (0.1-50 TeV) are expected to come from a wide variety of cosmic objects within and outside our galaxy. Studying this radiation yields valuable and unique information about the unusual astrophysical environment characterizing these sources, and also the intervening intergalactic space. At energies above 0.1 TeV, a typical gamma-ray source is generally too weak for direct detection by satellite based detectors necessitating the use of ground-based atmospheric Cerenkov detection technique. In atmospheric Cerenkov imaging telescopes one records the spatial distribution of the Cerenkov photons in the image plane of a light collector by using a close-packed array of photomultipliers. By examining the subtle differences in the shapes and orientation of the shower images, it becomes possible to reject more than 99.5% of the cosmic-ray generated background Cerenkov events. Gamma-ray events give rise to shower images with their major axes preferentially oriented towards the source position in the image plane. Apart from being narrow and compact in shape, these images have a cometary shape with their light distribution skewed towards their source position in the image plane and become more elongated as the impact parameter of the shower increases. On the other hand, hadronic events give rise to images that are, on average, broader and longer and are randomly oriented within the field of view of the camera. For each image, which is essentially elliptical in shape, Hillas parameters [1] are calculated to characterize its shape and orientation. The parameters are obtained using moment analysis and are defined as : LENGTH--The rms spread of light along the major axis of the image (a measure of the vertical development of the shower); WIDTH -- The rms spread of light along the minor axis of the image (a measure of the lateral development of the shower); DISTANCE-- The distance from the centroid of the image to the centre of the field of view;  $\alpha$  --The angle between the major

axis of the image and a line joining the centroid of the image to the position of the source in the focal plane; SIZE -- Sum of all the signals recorded in the clean Cerenkov image; FRAC2-- The degree of light concentration as determined from the ratio of the two largest PMT signals to sum of all signals (also referred to as Conc.).

Keeping in view the global developments in the field, the gamma-ray telescope TACTIC (TeV Atmospheric Cerenkov Telescope with Imaging Camera) [2] has been set up at Mt. Abu ( $24.6^{\circ}$  N,  $72.7^{\circ}$  E, 1300m asl), India. The telescope has been in operation since 2001 to study TeV gamma-ray emission from various celestial sources. The telescope uses a tessellated light collector of area  $\sim 9.5\text{m}^2$ , which is capable of tracking a celestial source across the sky. The light collector uses 34 front-face aluminum coated glass mirrors of 60cm diameter each with a focal length of  $\sim 400\text{cm}$ . The telescope deploys a 349-pixel imaging camera, with a uniform pixel size of  $\sim 0.31^{\circ}$  and  $\sim 5.9^{\circ} \times 5.9^{\circ}$  field of view, to record atmospheric Cerenkov events produced by an incoming cosmic-ray particle or gamma-ray with an energy above  $\sim 1\text{TeV}$ . With a  $5\sigma$  sensitivity of detecting the Crab Nebula in  $\sim 25$  hours of observation time, regular observations were taken on a number of potential gamma-ray sources (viz., Mrk 421, Mrk 501, 1ES2344+514, PSR 0355+54, ON 231, H1426 etc.) during the last 10 years.

The sensitivity of a Cerenkov imaging telescope is strongly dependent on the rejection of the cosmic-ray background events. The methods which have been used to achieve the segregation between the gamma-rays from the source and the background cosmic-rays, include methods like Supercuts/Dynamic Supercuts, Maximum likelihood classifier, Kernel methods, Fractals, Wavelets, Random Forest etc. The conventionally used Supercuts/Dynamic Supercuts method, though using several image parameters simultaneously, with some of them also being energy dependent, is still a one dimensional technique, in the sense that the parameters it uses for classification are treated separately and the possible correlations among the parameters are ignored. While the segregation potential of the neural network classifier has been investigated in the past with modest results, one of the main objectives of the thesis is to study the gamma / hadron segregation potential of various ANN algorithms, some of which are supposed to be more powerful in terms of better convergence and lower error compared to the commonly used Backpropagation algorithm. Furthermore, apart from detecting gamma-ray sources, one of the main aims of the Cerenkov imaging telescopes is also to reconstruct the energy spectra of the sources. In the thesis work, we have applied various artificial Neural Network (ANN) algorithms for improving the sensitivity of the TACTIC telescope. In addition, ANN is also used for determining the energy of the primary gamma-rays on the basis of their image SIZE, Distance and zenith angle.

A neural network is a parallel distributed information processing structure consisting of processing elements interconnected together with unidirectional signal channels called connections. Each processing element has a single output connection which branches into many collateral connections as desired. Depending upon the architecture in which the individual neurons are connected and the error minimization scheme adopted, there can be several possible ANN configurations. While algorithms like Standard backpropagation and the Resilient backpropagation come under the category of Local search algorithms, Conjugate Gradient methods, Levenberg-Marquardt algorithm, One Step Secant etc. belong to the category of Global search algorithm. Hybrid algorithm category constitutes models like Higher Order Neuron and Neuro Fuzzy systems etc. A brief summary of the work related to gamma / hadron segregation potential of

various ANN algorithms and ANN-based energy estimation procedure for determining the energy of the primary gamma-ray is given below.

The gamma/hadron segregation potential of various ANN algorithms was studied by applying them to the Monte Carlo simulated and actual observational data on the Crab Nebula. The network used in this work comprises 6 nodes in the input layer and one neuron in the output layer the value of which is used to categorize the output. We used the following six image parameters in the ANN-based gamma/hadron segregation methodology: Zenith angle, SIZE, LENGTH, WIDTH, DISTANCE and FRAC2. The results obtained suggest that Levenberg-Marquardt method outperforms all other methods in the ANN domain. Applying this ANN algorithm to  $\sim 101.44$  h of Crab Nebula data collected by the TACTIC telescope, during 2005 - 2006, yields an excess of  $\sim (1141 \pm 106)$  with a statistical significance of  $\sim 11.07 \sigma$ , as against an excess of  $\sim (928 \pm 100)$  with a statistical significance of  $\sim 9.40 \sigma$  obtained with Dynamic Supercuts selection methodology. The main advantage accruing from the ANN methodology is that it is more effective at higher energies and this has allowed us to re-determine the Crab Nebula energy spectrum in the energy range  $\sim 1$ -24 TeV. The results of this study have been published in [3].

A novel energy reconstruction procedure, based on the utilization of Artificial Neural Network (ANN), has also been developed for the TACTIC atmospheric Cerenkov imaging telescope. The procedure uses a 3:30:1 ANN configuration with Resilient backpropagation algorithm to estimate the energy of a gamma-ray like event on the basis of its image SIZE, DISTANCE and zenith angle. The new ANN-based energy reconstruction method, apart from yielding an energy resolution of  $\sim 26\%$ , which is comparable to that of other single element imaging telescopes, has the added advantage that it considers zenith angle dependence as well. The results of this study have been published in [4]. The ANN-based energy estimation procedure has also been used successfully for determining the time-averaged energy spectra of Mrk 501 and Mrk 421 which were found to be in high state by TACTIC on several occasions from 2006 to 2010 [5,6,7].

The contents of the thesis are organized in the following manner. In Chapter 1 of the thesis, we provide an introduction to the field of very high energy gamma-ray astronomy and Cerenkov imaging technique. Chapter 2 will cover the details regarding the hardware aspects of the TACTIC telescope. An overview of the ANN will be presented in Chapter 3. This chapter will also include the following aspects of the ANN: Working and architecture of Neural network, Training of ANN, Multilayered Feed forward Neural Networks, first and second order Learning algorithms. The comparative performance of the ANN algorithms by applying them to standard benchmarking problems like the IRIS data, XOR/N-bit parity and two-spiral problems will be presented in Chapter 4 [8]. In chapter 5 we will study the gamma/hadron segregation potential of various ANN algorithms by applying them to the Monte Carlo simulated data and actual TACTIC observation data on the Crab Nebula. The details of energy reconstruction procedure, based on the utilization of Artificial Neural Network (ANN), will be presented in Chapter 6. Finally in Chapter 7, we present our main conclusions.

**Publications in refereed Journals:**

- a. Published : 12  
b. Accepted : X  
c. Communicated : X  
d. Other Publications : X

Signature of Student



Date

Sept 10, 2013

**Doctoral Committee**

S. No	Name	Designation	Signature	Date
1	Prof. B.S. Acharya	Chairman		19-09-2013
2	Dr. A.K. Tickoo	Convener		Sept, 10, 2013
3	Dr. N.K. Gupta	Member		Sept. 16, 2013
4	Dr. A.K. Mohanty	Member		Set 10-2013
5	Mr. R. Koul	Member		10.09.2013

# Contents

<b>1</b>	<b>Introduction to Very High Energy Gamma-Ray Astronomy</b>	<b>2</b>
1.1	Gamma-Ray astronomy background . . . . .	2
1.1.1	Motivation for $\gamma$ -ray observations and terminology used . . .	3
1.2	Cosmic-Rays . . . . .	4
1.2.1	Air showers: Origin, development and sub-showers . . . . .	5
1.2.2	Differences between electromagnetic and hadronic sub-showers	9
1.3	Ground based gamma-ray astronomy . . . . .	10
1.3.1	Atmospheric Cherenkov technique . . . . .	11
1.3.2	Characteristics of Cherenkov radiation . . . . .	14
1.3.3	Extinction of Cherenkov light in the atmosphere . . . . .	16
1.4	Atmospheric imaging Cherenkov technique . . . . .	17
1.4.1	Detection principle . . . . .	17
1.4.2	Cherenkov image parameters . . . . .	20
1.4.3	Energy threshold and flux sensitivity . . . . .	22
1.5	Challenges of $\gamma$ -ray astronomy . . . . .	23



1.6	Milestones and present status of the field . . . . .	24
1.6.1	Galactic sources . . . . .	25
1.6.2	Extragalactic sources . . . . .	27
1.7	References . . . . .	29
<b>2</b>	<b>TACTIC telescope : Details and Performance</b>	<b>31</b>
2.1	Observatory site, mechanical assembly and drive control . . . . .	31
2.2	Light-collector design of the TACTIC telescope . . . . .	34
2.3	The imaging camera . . . . .	35
2.3.1	Photomultiplier tubes and light guides . . . . .	36
2.4	Backend signal processing electronics . . . . .	37
2.4.1	Trigger generation . . . . .	38
2.5	Data acquisition and control system . . . . .	38
2.6	Relative and absolute gain calibration scheme . . . . .	40
2.7	Monte Carlo simulations . . . . .	41
2.7.1	Comparison with real data . . . . .	43
2.8	Performance evaluation using Crab Nebula observations . . . . .	45
2.9	Sensitivity estimates . . . . .	47
2.10	References . . . . .	49
<b>3</b>	<b>Artificial Neural Network (ANN) Methodology</b>	<b>51</b>
3.1	Introduction . . . . .	51

3.1.1	Biological and Artificial Neural Networks (ANN)	52
3.2	Learning	55
3.3	Unsupervised methods	57
3.3.1	Kohonen Self Organizing Maps (SOMs)	57
3.3.2	Adaptive Resonance Theory (ART)	58
3.3.3	Hebbian learning	59
3.3.4	Drawbacks of unsupervised networks	60
3.4	Supervised methods	60
3.4.1	First order algorithms	60
3.4.2	Standard Backpropagation algorithm	61
3.4.3	Overview of training	61
3.4.4	Caveats of the standard backpropagation model	66
3.4.5	Resilient backpropagation	67
3.4.6	Quick-propagation algorithm	69
3.5	Second order algorithms	69
3.5.1	Conjugate gradient methods and its variants	69
3.5.2	Lavenberg-Marquardt algorithm	71
3.5.3	Radial basis functions	73
3.5.4	Simulated annealing	76
3.5.5	Secant Methods-One step secant	78
3.6	Hybrid algorithms	79

3.6.1	Higher order neurons . . . . .	79
3.6.2	Neuro Fuzzy systems . . . . .	81
3.6.3	Adaptive Neuro-Fuzzy Inference System (ANFIS) . . . . .	83
3.6.4	Compensatory Neuro-Fuzzy system . . . . .	86
3.6.5	Fuzzification and decision oriented fuzzy neuron . . . . .	86
3.7	Techniques to improve results in ANN . . . . .	87
3.7.1	Overfitting, generalization and size of networks . . . . .	88
3.7.2	Combining network outputs and pruning . . . . .	88
3.8	References . . . . .	89
<b>4</b>	<b>Comparative study of ANN algorithms</b>	<b>91</b>
4.1	Introduction . . . . .	91
4.2	Benchmarking of ANN algorithms . . . . .	92
4.2.1	IRIS problem . . . . .	93
4.2.2	XOR and N-bit parity problems . . . . .	95
4.2.3	Two-spiral problem . . . . .	97
4.3	Application of ANN algorithms to regression problems . . . . .	99
4.3.1	Approximation to $\cos(x)$ . . . . .	100
4.3.2	Approximation to a few special functions . . . . .	101
4.4	Summary . . . . .	104
4.5	References . . . . .	106

<b>5</b>	<b>ANN based gamma-hadron segregation for TACTIC telescope</b>	<b>108</b>
5.1	Introduction . . . . .	108
5.2	Brief description of some similar applications using ANN . . . . .	109
5.3	Statistical analysis of Cherenkov image parameters . . . . .	111
5.3.1	Statistical analysis of various parameters for selecting the optimal features . . . . .	111
5.4	ANN methodology . . . . .	116
5.5	Gamma/hadron separation using ANN . . . . .	117
5.5.1	Preparation of training, testing and validation data . . . . .	117
5.5.2	ANN training and optimizing the hidden layer nodes . . . . .	118
5.5.3	Testing and validating Lavenberg-Marquardt method . . . . .	123
5.6	Determination of optimum ANN cut value . . . . .	123
5.7	Application of the ANN methodology to the Crab Nebula data collected with the TACTIC telescope . . . . .	127
5.8	Comparison of Dynamic Supercuts and ANN analysis methods . . . . .	131
5.9	Summary . . . . .	135
5.10	References . . . . .	135
<b>6</b>	<b>Energy reconstruction using ANN</b>	<b>138</b>
6.1	Introduction . . . . .	138
6.1.1	Factors governing the energy reconstruction . . . . .	139
6.2	Monte Carlo simulations for energy reconstruction of $\gamma$ -rays . . . . .	141

6.3	Conventional energy reconstruction methods . . . . .	143
6.3.1	Parameterized fit with DISTANCE and SIZE as variables . .	143
6.3.2	Parameterized fit with DISTANCE, SIZE and zenith angle as variables . . . . .	145
6.3.3	Look-up table method using interpolation in 3 dimensions .	147
6.4	Energy reconstruction using ANN . . . . .	148
6.4.1	Training of the network . . . . .	148
6.4.2	Testing and validation of the ANN . . . . .	150
6.5	Energy spectrum of the Crab Nebula as measured by the TACTIC telescope . . . . .	153
6.6	Summary . . . . .	154
6.7	References . . . . .	156
<b>7</b>	<b>Conclusions and future work</b>	<b>158</b>
7.1	Brief summary of the work presented . . . . .	158
7.2	Multivariate background rejection and the future of ANN . . . . .	162
7.3	TACTIC Telescope upgradation and the road ahead . . . . .	163
7.4	Other possible applications of ANN . . . . .	165
7.4.1	Image cleaning with ANN . . . . .	165
7.4.2	Application of ANN to raw images . . . . .	165
7.5	New Experiments . . . . .	166
7.5.1	MACE Telescope . . . . .	166

7.5.2 Cherenkov Telescope Array (CTA) . . . . .	167
7.6 New insights to VHE $\gamma$ -ray observations . . . . .	168
7.7 References . . . . .	169
<b>Appendices</b>	<b>170</b>
<b>A Gamma-ray production mechanisms</b>	<b>171</b>
<b>B Interaction of <math>\gamma</math>-rays</b>	<b>176</b>



# Chapter 1

## Introduction to Very High Energy Gamma-Ray Astronomy

### 1.1 Gamma-Ray astronomy background

The presence of certain ionizing radiation on the surface of Earth was initially predicted by Rutherford [1.1] after observing the discharge of a gold leaf electroscope. This radiation was postulated to be of terrestrial origin until balloon experiments of Goeckel [1.2], observed that radiation kept decreasing upto an altitude of  $\sim 1$  km, beyond which it increased. Victor Hess approached this mystery by himself measuring the radiation upto an altitude of  $\sim 5$  km aloft in a balloon. He concluded that there was a radiation penetrating the atmosphere from the outer space which was fittingly named as the ‘cosmic-radiation’. These cosmic-rays are now known to be consisting of variety of highly energetic particles. Over the inter-stellar distances these charged particles are deflected by the magnetic fields so that by the time they reach the ground level, all information about their direction is lost. The neutral component of this cosmic radiation consisting of neutrinos, neutrons and  $\gamma$ -rays however retain their direction information and are thus of vital astrophysical significance. However, neutrinos have an extremely small interaction cross section with matter and are thus difficult to detect. The life time of neutrons is too short to provide an appreciable detectable flux.  $\gamma$ -rays on the other hand, can travel over inter-galactic distances without interacting. More importantly, they are destroyed rather than scattered by interactions, thus any survived  $\gamma$ -ray will have travelled directly from the source. The study of  $\gamma$ -rays can thus provide invaluable information about cosmic-ray sources. It also gives us an insight into the high energy processes which occur in these extreme astrophysical environments.

The field of  $\gamma$ -ray astronomy and astrophysics has witnessed dramatic development on the observational and experimental fronts in the recent past. New generation experiments have been carried out both in space and on ground, following which our perspective of the high energy universe has been substantially enriched,



both in content and overall range. In the lower energy bracket of the  $\gamma$ -ray window (0.1 to 100 GeV), accessible to the satellite based experiments like the Fermi LAT launched in 2008, the  $\gamma$ -ray map comprises a large number of galactic and extragalactic sources ( $> 2000$ ), including several with no known counterparts in other regions of the electromagnetic (em) spectrum. Likewise, high energy  $\gamma$ -rays (100 GeV-10's TeV), accessible to various ground-based Cherenkov imaging telescopes, have enabled us to unequivocally detect various  $\gamma$ -ray source populations and study their temporal and spatial characteristics at unmatched precision levels. All these reassuring recent developments on the observational front have made the  $\gamma$ -ray astronomy field very exciting.

### 1.1.1 Motivation for $\gamma$ -ray observations and terminology used

Many energetic cosmic processes produce  $\gamma$ -ray radiation over a wide range of energies. Cosmic  $\gamma$ -rays span about 10 decades of energy from  $\sim 10^6$  to  $\sim 10^{16}$  eV, which is wider than the combined range of emission of em radiation from radio-waves to X-rays. Gamma-rays provide the best window for the study of non-thermal universe and an excellent tool for probing fundamental physics beyond the reach of terrestrial accelerators. Apart from detecting new  $\gamma$ -ray sources, one of the main aims of  $\gamma$ -ray telescopes is to reconstruct the energy spectra of the sources. A study of the resulting spectral energy distributions can yield valuable information about the underlying  $\gamma$ -ray production mechanisms and unusual astrophysical environment characterizing these sources. By studying the  $\gamma$ -ray emission produced by Supernova Remnants (SNR) we can determine if it is produced by electrons or protons and establish whether or not SNRs are the main sites of Galactic cosmic-ray acceleration. Gamma-ray observation is the most sensitive method to measure the flux of cosmic electrons in the energy range 20 GeV-5 TeV because it is claimed to contain a component which cannot be accounted for by the conventional models of cosmic-ray propagation and interaction in galaxy.

By studying  $\gamma$ -ray emission from sources at cosmological distances we can indirectly measure the extragalactic background light (EBL). This is the light emitted by all extragalactic sources over the history of the universe. Measurement of this radiation is important as it provides constraints to models of star formation, galaxy evolution and cosmology. In addition, differences in the observed energy spectrum of several active galactic nuclei can also be used to study absorption effects at the source or in the intergalactic medium due to the interaction of  $\gamma$ -rays with the extragalactic background photons [1.3-1.4].

The term  $\gamma$ -ray is generally used to denote electromagnetic photons with energies above several hundred thousand electron volts (eV). Therefore  $\gamma$ -rays occupy an extraordinarily broad and unbounded range of em spectrum. The term  $\gamma$ -ray astronomy is too broad and a subdivision is clearly needed. The most logical subdivision

was suggested by Weekes in his review paper [1.5]. He proposed a classification scheme based primarily on the range of energies over which a detection technique is used. However, since then, the classification has been modified to account for certain more divisions and what we report here is an updated nomenclature for  $\gamma$ -ray energy range to reflect the current status. The term high energy [HE] is used to include  $\gamma$ -rays from 30MeV-30GeV, a region which is explored by satellite based detectors. The region between 30GeV-30TeV, called the very high energy region [VHE] is narrower than what was chosen in [1.5] originally. The ultra-high energy (UHE) region from 30TeV-30PeV, describes the range studied through the air-shower particle detector arrays. Energies above this range i.e  $\geq 30\text{PeV}$  is defined as the extremely high energy range (EHE) which can be studied by the air shower/fluorescence techniques. Table 1.1 summarises the various ranges along with their detection techniques.

<b>Energy Range</b>	<b>Nomenclature</b>	<b>Detection Method</b>
10 – 30 MeV	<i>Medium</i>	<i>Satellite based Compton telescope</i>
30 MeV – 30 GeV	<i>High(HE)</i>	<i>Satellite based tracking detector</i>
30 GeV – 30 TeV	<i>Very high Energy(VHE)</i>	<i>Ground based atmospheric Cherenkov detector</i>
30 TeV – 30 PeV	<i>Ultra high Energy(UHE)</i>	<i>Ground based Particle detector</i>
$\geq 30\text{PeV}$	<i>Extremely high Energy (EHE)</i>	<i>Ground based Particle detector</i>

Table 1.1:  $\gamma$ -ray classification on the basis of energy.

## 1.2 Cosmic-Rays

Until HE accelerators came into being in 1950's, experiments in the  $\gamma$ -ray domain were conducted with cosmic-rays and particles like muons, pions etc were actually discovered in cosmic-ray experiments. It was thus natural that while studying  $\gamma$ -ray astronomy the questions about origin and composition of cosmic-rays became of prime importance.

Cosmic radiation consists of subatomic particles that possess a wide range of energies. The rate at which cosmic-rays hit the Earth varies with their energy. Low energy cosmic-rays are plentiful (many thousand  $\text{m}^{-2} \text{s}^{-1}$ ). Though some of

the cosmic-rays originate from the Sun, most of them come from sources outside the solar system. The composition of cosmic radiation is given in Table 1.2. It is

<b>Particle</b>	<b>Ratio</b>
<i>Proton</i>	85%
$\alpha$ – <i>particle</i>	12%
<i>Electron</i>	2%
<i>Heavy Nuclei</i>	$\leq 1\%$
<i>Photon(<math>\gamma</math> – ray)</i>	$\leq 0.2\%$

Table 1.2: Composition of Cosmic-Rays

clear from the table that a small but important contribution to the composition of cosmic-rays comes from neutral particles, namely photons.

### 1.2.1 Air showers: Origin, development and sub-showers

As a high energy cosmic-ray particle enters the Earth’s atmosphere it creates an extensive air shower (EAS) by interaction with the particles present in the atmosphere. Strong electromagnetic interactions lead to development of air showers. In these interactions, secondary particles are created which in turn interact with atmospheric particles and create new particles via a chain reaction. Thus a cascade of particles is generated by these interactions which is called an air shower. Photons interact mainly via electromagnetic interaction, leading to electron-positron pair production and then by means of bremsstrahlung process (see Appendix), it eventually creates the electron-photon showers. Protons, on the other hand, interact with nuclei of the air molecules via strong interaction leading to hadronic cascades. In these interactions the particles loose energy till the average energy per particle drops below the threshold critical energy ( $\sim 86$  MeV). At this point the shower reaches its maximum size called ‘shower maximum’ beyond which particles suffer energy loss due to ionization, as a result the number of shower particles decay exponentially and finally the shower dies out.

#### Electromagnetic showers

In case of an electromagnetic cascade, three processes dominate the longitudinal development, the bremsstrahlung, the pair production processes and the ionization processes which causes the shower to disintegrate. A discussion on these production processes is beyond the scope of this thesis, however a short introduction to these has been provided in the appendix for the sake of completeness.

High energy photons bombarding the atmosphere, convert into a electron positron pair in the coulomb field of atomic nuclei (pair production process). The electrons

and positrons are deflected by the nuclei and emit photons via the bremsstrahlung process. Thus both the processes set an avalanche of particles in tandem which results in an electromagnetic shower. The shower growth continues till energy of the particles drops below some critical energy ( $E_c$ ) where the energy loss of electrons by ionization of the air molecules becomes dominant. At this stage the shower reaches its maximum and no new particles can be created further. Accordingly the shower is dissipated by the ionization of the medium.

The characteristic length scale of these interaction processes is given by the mean free path length. In case of the bremsstrahlung process, this is referred to as the radiation length  $X_0$  and is measured in  $\text{g cm}^{-2}$ . It can be understood as the mean distance after which a high energy electron loses all but  $\left(\frac{1}{e}\right)$  of its initial energy  $E_0$  by the bremsstrahlung process. This can be mathematically represented by:

$$\left(-\frac{dE}{dx}\right) = \frac{E_0}{X_0} \quad (1.1)$$

where  $\frac{dE}{dx}$  is the energy loss of an electron due to bremsstrahlung after travelling a distance  $x$  in matter. Integrating the above expression leads to

$$E(x) = E_0 \cdot e^{-\frac{x}{X_0}} \quad (1.2)$$

In air, the radiation length  $X_0$  is  $\sim 37.2 \text{g cm}^{-2}$  and the corresponding e-folding distance for pair production by high energy photons, called the conversion length  $X_c$ , is  $\sim \frac{9}{7} \cdot X_0$ .

The predominant process which determines the lateral shower development is the multiple coulomb scattering. A charged particle traversing air is deflected through small angles by the air molecules present in the atmosphere. Most of this deflection stems from coulomb scattering of electrons or photons against nuclei.

Bremsstrahlung and pair production contribute to the lateral spread of the secondary particles with respect to the shower axis as well. An electron undergoing bremsstrahlung radiates a photon in a cone in forward direction with an average opening angle  $\theta = \frac{1}{\gamma}$ , where  $\gamma$  is the Lorentz factor. Thus for high energy electrons, the directional divergence from the shower axis originating from bremsstrahlung is very small and can be neglected.

The basic properties of the development of the electromagnetic shower can be seen on the basis of the model proposed by Bethe and Hitler [1.6]. It is assumed that both the radiation and conversion lengths are equal, namely  $X_0$ , and the incident energy  $E_0$  is divided equally between the secondary particles. Fig 1.1 illustrates the assumptions. After 'n' radiation lengths, the shower consists of  $2^n$  particles each having an energy of  $E(n) = E_0 \cdot 2^{-n}$ . The depth of the shower maximum in the atmosphere,  $t_{max}$ , is then given by the expression:

$$t_{max} = \frac{\ln \frac{E_0}{E_c}}{\ln 2} \cdot X_0 \quad (1.3)$$

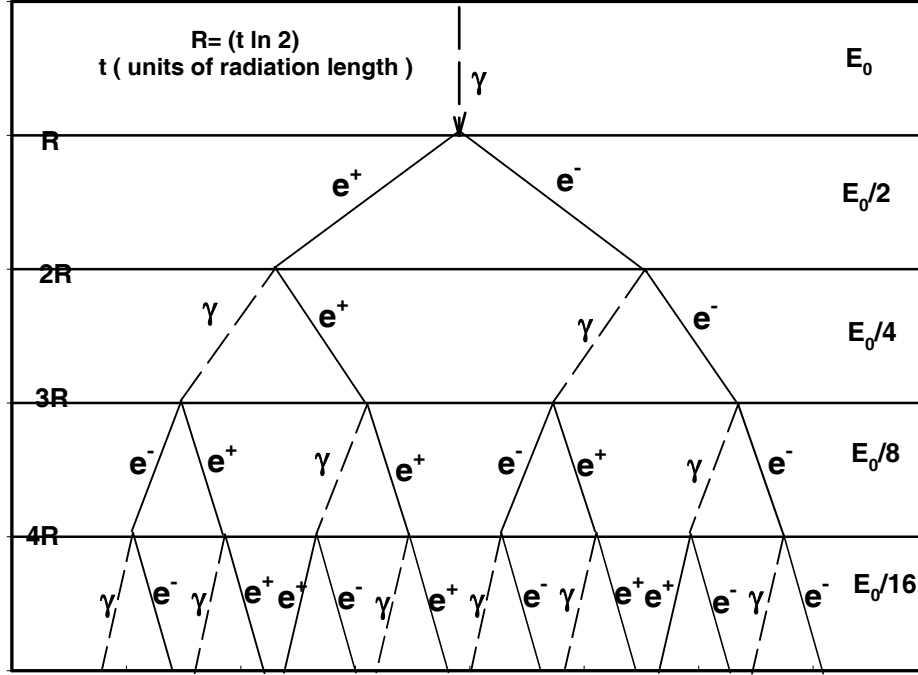


Figure 1.1: A simplified model of Electromagnetic cascade.

From above one can obtain the number of particles in the shower maximum to be

$$N_{max} = \frac{E_0}{E_c} \quad (1.4)$$

where  $2/3^{rd}$  of the particles are electrons and positrons and the rest are photons.

This simple model predicts an exponential increase in the number of particles in the initial phase of the shower development. The maximum amount of particles is proportional to the energy of the primary particle and the depth of the shower maximum grows logarithmically with the primary energy.

### Hadronic showers

The shower development for hadrons (protons) hitting the Earth's atmosphere differs slightly from the shower development for photons and electrons discussed above. Hadrons interact only through the strong interaction with the nuclei of the air molecules. These interactions are more complex as both hadronic as well as the electromagnetic showers are involved. In the first interaction of the hadron with a particle in the Earth's atmosphere, strong interaction is dominating. The interaction produces not only fragments of the target nucleus, like pions and kaons, but

nucleons (neutrons and protons) and hyperons are also created. These secondary particles, are energetic enough to interact strongly, and hence a hadronic cascade is created.

Neutral pions ( $\pi^0$ ) which carry about one-third of energy of all generated pions ( $\pi^+$ ,  $\pi^-$ ) play an important role in hadronic cascade. The neutral pions decay almost instantly after a mean lifetime of  $\sim 10^{-17}$  seconds into two photons as depicted in Fig 1.2. The photons generate electromagnetic subshowers in the hadronic cascade

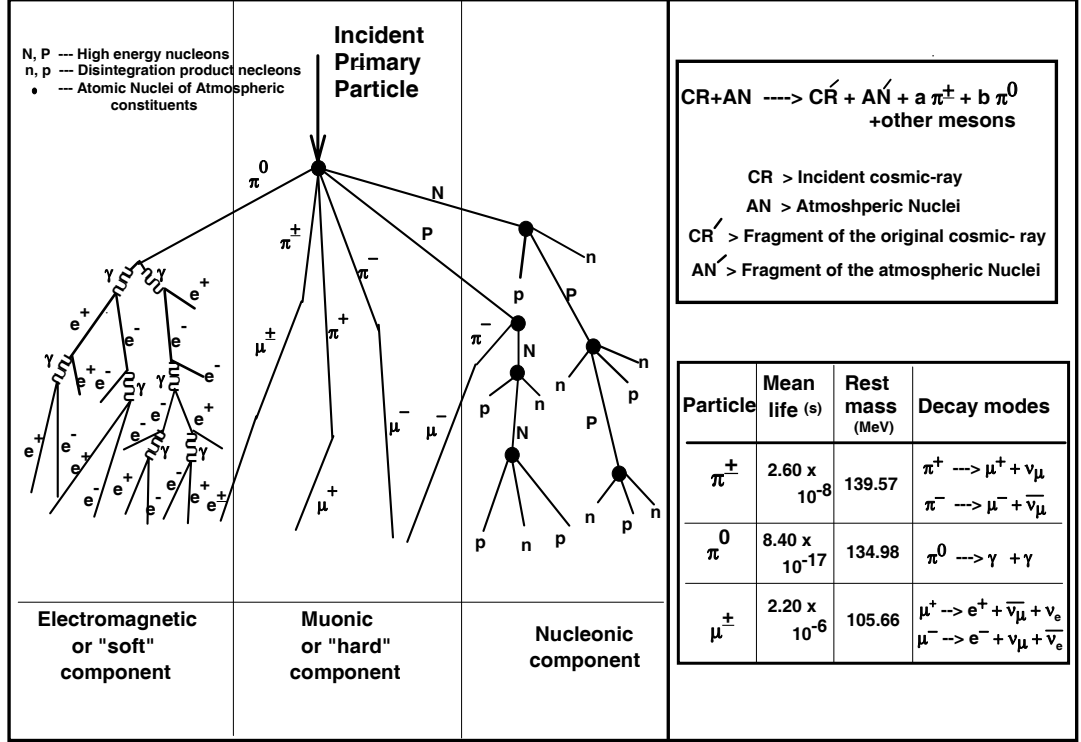


Figure 1.2: Development of a hadronic cascade. The principal modes of energy propagation of cosmic-ray cascade are also shown in the figure.

as described in the previous section. The charged pions ( $\pi^\pm$ ) have a much longer lifetime and decay after  $\sim 10^{-8}$  seconds mainly into muons ( $\mu^\pm$ ) which in turn decay into electrons and neutrinos. The longer lifetime of the charged pions results in an increasing probability of new interactions with the nuclei of the air molecules during their lifetime. These interactions give rise to hadronic subshowers in the hadronic cascade. The different types of subshowers of the hadronic shower can also be seen in Fig 1.2.

### 1.2.2 Differences between electromagnetic and hadronic sub-showers

The interaction length of hadrons is  $\sim 85 \text{ g cm}^{-2}$  and is larger as compared to  $35 \text{ g cm}^{-2}$  for photons. Therefore the hadrons penetrate deeper into the atmosphere, leading to a larger shower depth. In these hadronic interactions a large part of the shower energy is lost in the generation of new particles like muons ( $\mu^\pm$ ), mesons, and other secondary hadrons. The lateral development of electromagnetic showers on the other hand is determined by elastic coulomb scattering of electrons, thus for the TeV photons, the mean scattering angle is very small (see Fig 1.3), which means that the lateral spread of the electromagnetic shower is smaller. The secondary particles participating in the hadronic shower, on the other hand, receive a higher transverse momentum by inelastic scattering at extended target particles like the nuclei of the air molecules. Complex multiparticle processes are involved in the

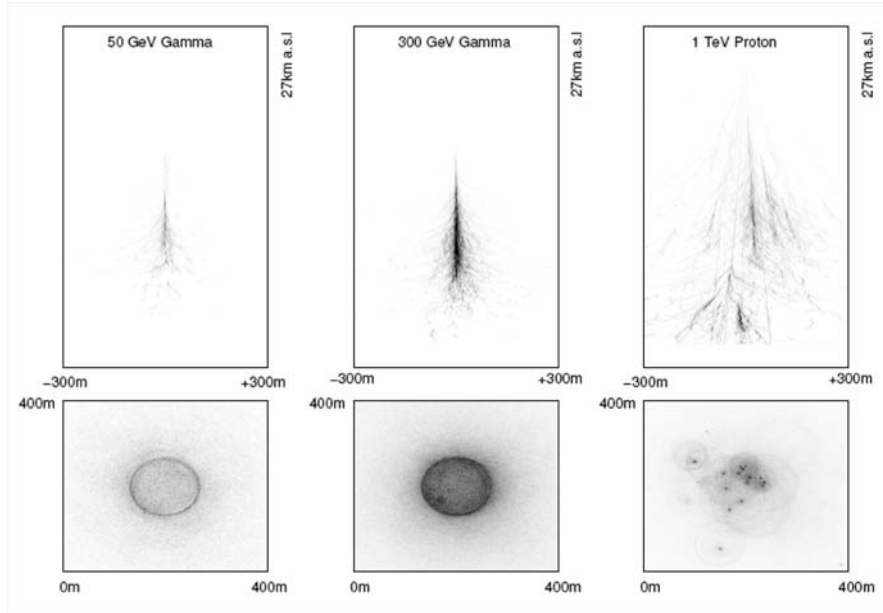


Figure 1.3: The vertical development and top view of the Cherenkov pool, for a 50GeV  $\gamma$ -ray, 300GeV  $\gamma$ -ray and 1TeV proton initiated simulation showers. Adapted from [www.ast.leeds.ac.uk/fs/showerimages.html](http://www.ast.leeds.ac.uk/fs/showerimages.html).

development of the hadronic shower in contrast to mainly three-particle processes like bremsstrahlung and pair production in the electromagnetic shower. Therefore, the hadronic shower is less regular, has larger fluctuations, and contains electro-

magnetic sub-showers that are created by neutral pion decays (Fig 1.3). It is these differences between the two event species which ground based  $\gamma$ -ray astronomy seeks to exploit [1.7].

### 1.3 Ground based gamma-ray astronomy

Several attempts were made in the past by various groups to detect  $\gamma$ -ray sources using atmospheric Cherenkov technique. Generally speaking, the detection systems deployed in the past were not sufficiently sensitive and these experiments met with a limited success only. Ground based VHE  $\gamma$ -ray astronomy came into prominence with the first ever detection of TeV  $\gamma$ -ray source, Crab Nebula by the 10 m diameter Whipple imaging Cherenkov telescope. Since then, VHE  $\gamma$ -ray astronomy, in the energy range  $\sim 100\text{GeV}-30\text{TeV}$ , has matured into an exciting field of research.

While as, HE  $\gamma$ -rays can be detected from space by space telescopes like the Large Area Telescope (LAT) on board the Fermi satellite observatory, VHE  $\gamma$ -rays can be detected from Earth by ground based imaging telescopes like MAGIC, HESS, VERITAS, Whipple, TACTIC etc. The space based Fermi-LAT consists of an anti-coincidence shield plus a tracker and a calorimeter, which allows a nearly background free detection of  $\gamma$ -rays in the energy bracket  $\sim 30\text{MeV}-100\text{GeV}$ . It has a wide field of view of  $\sim 0.8\pi$  steradians and a duty cycle of  $\sim 100\%$  [1.8]. It works in survey mode covering the full sky in every 3 hours. In the five years of its operation, out of  $\sim 2000$  sources detected by it, many of these are unidentified sources. On the other hand, ground based Imaging Atmospheric Cherenkov telescopes (IACTs) are sensitive to a higher energy range between  $\sim 100\text{GeV}$  to  $30\text{TeV}$ . The typical FOV of IACT is only few degrees ( $3-5^\circ$  typical) and they operate in pointing mode with a modest duty cycle of at best 10-15%, as the observations on  $\gamma$ -ray sources require moonless and cloud free nights. These telescopes work on the principle of capturing the images of Cherenkov light produced by electromagnetic showers initiated by  $\gamma$ -rays in the Earth's atmosphere. The main background affecting the observations of  $\gamma$ -rays using the technique is the overwhelming flux of charged cosmic-rays (CRs) which is at least 100 times more abundant than the  $\gamma$ -rays. This background is reduced during offline analysis by exploiting certain subtle differences between the two event species. Using this technique, over a dozen sources were detected in 1990's with IACTs. The first exploratory instruments were replaced in 2000's by more sophisticated instruments employing fast electronics and better designs which revolutionized the field. Gamma-ray source catalogue now has  $> 160$  sources and many new populations have been established as the  $\gamma$ -ray emitters which include SNRs, pulsar wind nebulae, radio galaxies and binary systems.

At still higher energies ( $\geq 50\text{ TeV}$ ), non-imaging air shower detectors identify showers initiated by photons using different experimental techniques. Unlike the IACTs, the non-imaging experiments cover only the northern hemisphere and have a limited angular resolution ( $\geq \pm 1^\circ$ ). Despite their better duty cycle ( $\geq 90\%$ )



and large field of view ( $\geq 2$  sr) the obtained sensitivity is far lower than those of IACTs.

The success of ground based  $\gamma$ -ray astronomy over the last two decades is mainly due to the advent of the Cherenkov imaging technique which allows efficient separation of photon induced showers from the huge deluge of the hadron background. In the following section we shall discuss the technique in detail.

### 1.3.1 Atmospheric Cherenkov technique

#### Cherenkov radiation

Production of light via the Cherenkov effect was initially proposed by Heaviside [1.9], and arguably, Marie Curie was the first to notice the bluish white light characteristic of Cherenkov radiation emanating from glass vials containing concentrated radium solution. Since this discovery was made in 1910 around the discovery of nuclear radiation, Cherenkov light has been observed in various dielectric media in the vicinity of radioactive materials [1.10]. The prerequisite properties of the particles to allow production of this radiation is that apart from being charged, they have to be very energetic also. Cherenkov radiation can thus be viewed as a shock wave propagating within a dielectric medium and exciting the medium to emit the radiation. The first systematic study of the effect was made by Mallet who was able to show that the spectrum was continuous, which distinguishes it from the spectra due to fluorescence. Cherenkov, then conducted a series of experiments between 1934 to 1938 [1.11] during which he demonstrated that the production site of radiation was modified by the presence of magnetic fields, i.e, Cherenkov emission in the vicinity of an uncharged ionizing radiation is due to the secondary particles. The threshold energy for the emission of Cherenkov light depends on the mass of the particle  $m_0$  and the refractive index  $n$  of the medium.

$$E_{min} = \frac{m_0 c^2}{\sqrt{1 - n^{-2}}} \quad (1.5)$$

This implies that light particles like the electrons and positrons dominate Cherenkov emission in the air showers. As the density of air changes continuously with altitude and also with the refractive index, the threshold energy as well as the emission angle depend on the altitude in the atmosphere.

#### Cherenkov effect

Consider an electron moving with velocity  $v < (\frac{c}{n})$  in a transparent dielectric medium of refractive index  $n$  as shown in Fig 1.4a. Circles shown represent the individual atoms of the dielectric medium. Normally these are expected to be roughly symmetrical and undistorted. However in the region close to the slowly

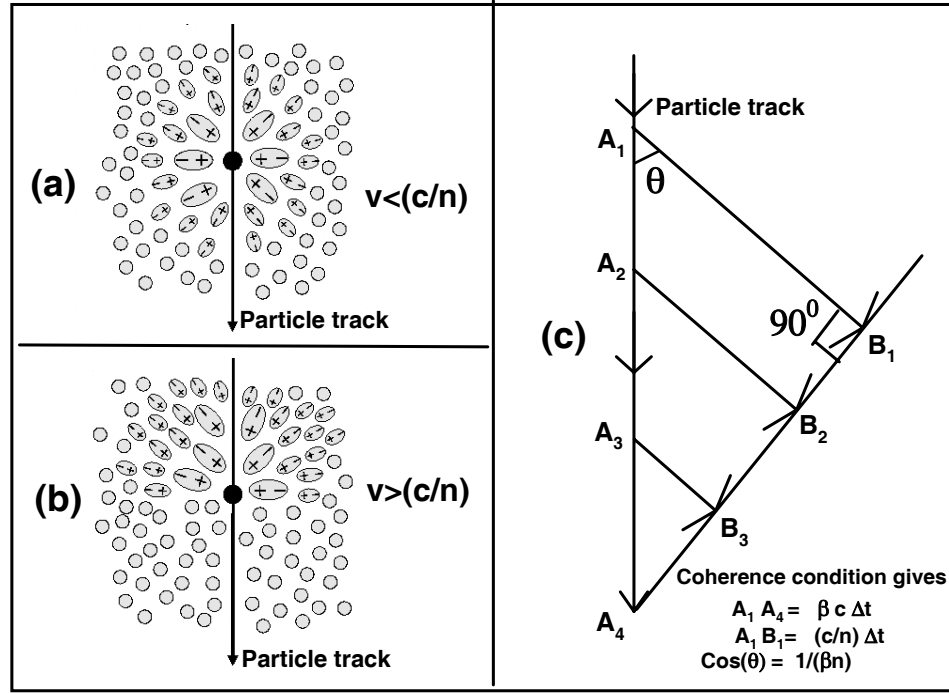


Figure 1.4: Polarization of medium by passage of a charged particle. (a) Case  $v < (\frac{c}{n})$ ; wavelets not in phase and interfere destructively (b) Case  $v > (\frac{c}{n})$ ; wavelets coherent and lead to Cherenkov emission. (c) Huygens construction for constructive interference of waves.

moving electron, the electric field of the electron polarizes the atoms of the medium isotropically, and when the electron moves to another point the distorted atoms return to their normal shape. However there is a complete symmetry of the polarization field surrounding the electron as well as along the trajectory so that as the particle passes through the medium, the net em field created at large distances vanishes.

On the other hand if  $v > (\frac{c}{n})$ , the induced polarization vector, instead of pointing towards the instantaneous position of the electron points towards its earlier position. This phenomenon which is essentially a retarding effect manifests itself in a flip-over of the axial component of the polarization vector because the electron location itself becomes a source of radiation with the pulse propagating upwards. This is shown in Fig 1.4b which shows that while the azimuthal plane symmetry is preserved, however along the axis, the resultant dipole is apparent even at large distances from the electron track. Therefore, it is possible for wavelets from all positions to be in phase so that a resultant field exists even at a distant point. This resultant radiation is referred to as the Cherenkov radiation. Using Huygens construction as

depicted in Fig 1.4c, this radiation is seen at a particular angle  $\theta$  with respect to the particle track. It is important to note here that what is demonstrated is only a 2 dimensional case. In reality, the light originating from each element of the track is actually propagated along the surface of a cone whose axis coincides with the track of the particle and whose semi-vertical angle is  $\theta$ . The condition of coherence implies that the time taken by the light to cover a distance  $A_1B_1$  equals the time it takes for the electron to cover distance  $A_1A_4$ . Thus ,

$$\frac{A_1B_1}{c/n} = \frac{A_1A_4}{v} \quad (1.6)$$

or

$$\cos\theta = \frac{1}{\beta n} \quad (1.7)$$

where  $\beta = \frac{v}{c}$ . It follows from the above equation that:

(a) For a medium of a given refractive index, there is a threshold velocity  $\beta_{min}=1/n$ , below which no radiation is emitted. At this critical velocity the direction of the radiation coincides with that of the particle.

(b) For an ultra-relativistic particle (i.e.  $\beta \sim 1$ ) there is a maximum angle of emission given by:

$$\theta_{max} = \cos^{-1} \left( \frac{1}{n} \right) \quad (1.8)$$

(c) The radiation occurs mainly in the visible and near visible regions of the electromagnetic spectrum for which  $n > 1$ . Emission in the X-ray region is impossible, because  $n$  turns out to be then less 1 in this case. There are two more conditions which need to be fulfilled to achieve coherence. First, the track length of the particle in the medium should be large compared with the wavelength ( $\lambda$ ) of the radiation in question; otherwise, diffraction effects dominate. Secondly, the velocity of the particle must be constant during its passage through the medium, or, to be more specific, the difference in the time for the particle to traverse successive distance  $\lambda$  should be small compared with the period  $\lambda/c$  of the emitted wave. These conditions are easily satisfied in the visible region, even in the presence of energy loss due to ionization ( $\sim 2$  MeV per g cm<sup>-2</sup>). It may be worth pointing out here that Cherenkov radiation phenomenon is completely different from recombination and excitation radiation (radiation associated with the ionization caused by the particle). It also differs fundamentally from bremsstrahlung radiation which arises due to the deceleration of the electrons mainly in the coulomb field of a nucleus. In this sense, bremsstrahlung radiation is akin to the radiation emitted by any accelerated charged particle. On the contrary, Cherenkov radiation is essentially a cooperative effect arising from the constituents of the medium through which the electron moves with essentially uniform velocity whose magnitude exceeds  $c/n$ . Furthermore, the total energy radiated by an electron in the bremsstrahlung process

is considerably greater than that emitted by Cherenkov radiation, while the very different spectral distributions in the two cases result in the intensity of Cherenkov radiation exceeding the bremsstrahlung by a large factor in the visible region.

So far, we have considered the radiation produced by the passage of a point charge through a dielectric medium, i.e. the normal Cherenkov effect. However, Cherenkov radiation is also produced when an electric or magnetic dipole passes through the medium, provided of course, its velocity is above a certain threshold. The radiation mechanism is the same as for the normal Cherenkov effect, though here intensity is much lower (by a factor of  $10^5$  to  $10^6$  than the normal Cherenkov effect) due to the fact that the field associated with a dipole falls off rapidly with distance, as compared to that for a point charge. Spectral distribution of Cherenkov radiation in this case is  $(\frac{1}{\lambda^5})d\lambda$  as against the  $(\frac{1}{\lambda^3})d\lambda$ . A detailed classical treatment of the problem of Cherenkov emission from moving dipoles has been discussed by Balazs [1.12].

### 1.3.2 Characteristics of Cherenkov radiation

#### Threshold kinetic energy and Cherenkov emission angle

Under the assumption of an exponential atmosphere, the refractive index  $n$  at a height  $h$  can be written in the following manner:

$$n = \left( 1 + n_0 \exp\left(\frac{-h}{h_0}\right) \right) \quad (1.9)$$

where  $h_0 = 7.1$  kms is the scale height of the atmosphere and  $n_0 = 2.7 \times 10^{-4}$ . It can be easily found out from Equation (1.7) that the expression connecting kinetic energy of the particle ( $T$ ), its rest mass ( $=m_0c^2$ ) and Cherenkov emission angle ( $\theta$ ) is given by:

$$T = \left( \frac{n \cos \theta}{\sqrt{n^2 \cos^2 \theta - 1}} - 1 \right) m_0 c^2 \quad (1.10)$$

Thus at the threshold ( $\theta = 0^\circ$ ), the minimum kinetic energy,  $T_{min}$  which a particle should have so that Cherenkov light can be produced, is given by :

$$T_{min} = \left( \frac{n}{\sqrt{n^2 - 1}} - 1 \right) m_0 c^2 \quad (1.11)$$

For electrons of rest mass energy  $\sim 0.511$ MeV,  $T_{min}$  is  $\sim 21$ MeV at sea level. Variation of  $T_{min}$  as a function of the atmospheric height for different particles is shown in Fig 1.5a. The expression for Cherenkov emission angle (for  $\beta \sim 1$ ) as a function of height 'h' in the atmosphere also follows from equation (1.7) and is given by

$$\theta = \sqrt{2n_0} \exp\left(\frac{-h}{2h_0}\right) \quad (1.12)$$

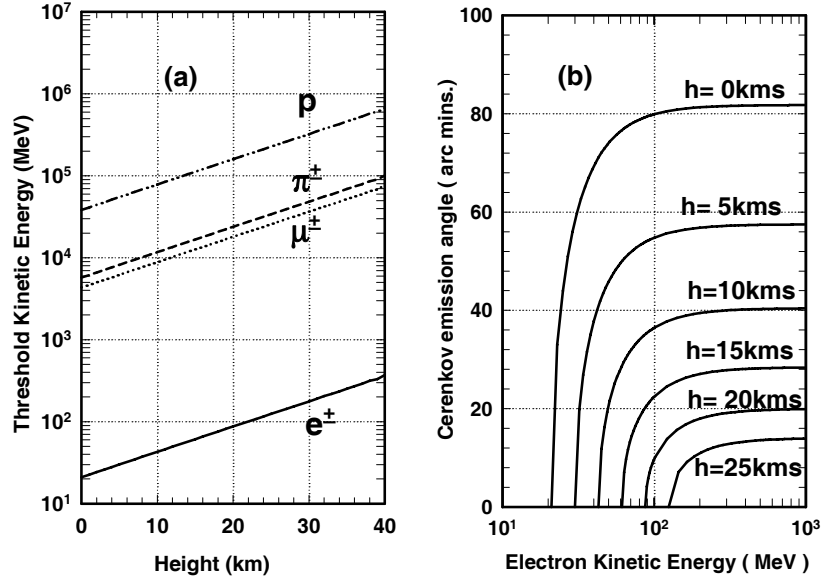


Figure 1.5: (a) Threshold kinetic energy for particles as a function of altitude for producing Cherenkov light. (b) Cherenkov angle and minimum kinetic energy required for electrons at different altitudes in the atmosphere.

Variation of  $\theta$  and  $T$  at different heights can be seen in Fig 1.5b.

### Cherenkov pool radius and time duration of the flash

Consider a single fast electron moving vertically downwards in the Earth's atmosphere. Neglecting its scattering and slowing down, let a Cherenkov light pulse generated at a height  $h$  and at an emission angle of  $\theta$ , arrive at the ground at a point  $r$  away from where the electron would have normally hit the ground. Since  $r$ ,  $h$  and  $\theta$ , are connected by  $r = h \tan \theta$ , we have

$$r(h) = \sqrt{2n_0 h} \exp\left(\frac{-h}{2h_0}\right) \quad (1.13)$$

The function  $r(h)$  has a maximum value  $r_{max}$  at  $h = h_{max} = 2h_0$  and equals  $\sim 126$  m. In other words, it means that Cherenkov photons produced in the region  $h_0 < h < 2h_0$  reach the ground level with their  $r$  values less than  $r_{max}$ . The atmospheric pressure at this typical height is  $140 \text{ g cm}^{-2}$ , with  $\eta = 4 \times 10^{-5}$ ,  $\theta = 9 \times 10^{-5}$  radians and  $T_{min} = 56 \text{ MeV}$ . In order to calculate the duration of a typical Cherenkov flash, let us consider a charged particle moving vertically downwards through the atmosphere. Let  $h_2$  and  $h_1$  (where  $h_2 > h_1$ ) be the heights at which the particle produces

Cherenkov light in the form of two photon bunches. Assuming a simplified situation where the Cherenkov light is emitted along the track of the particle itself, the duration of the Cherenkov flash can be calculated in the following manner:

Time recorded at the ground when the first photon bunch (emitted at  $t=0$  from a height  $h_2$ ) reaches ground is equal to  $\frac{1}{c} \int_0^{h_2} n \, dh$ . After a time  $(h_2 - h_1)/c$ , the particle moving with  $\beta \sim 1$  reaches height  $h_1$  and again emits a bunch of photons which strike the ground at time  $(h_2 - h_1)/c + 1/c \int_0^{h_1} n \, dh$ . Since the particle moves faster than the light flash in air, the second bunch arrives at ground earlier and the difference in the time interval at ground  $\delta t$  is given by :

$$\delta t = \frac{1}{c} \int_0^{h_2} n \, dh - \left( \left( \frac{h_2 - h_1}{c} \right) + \frac{1}{c} \int_0^{h_1} n \, dh \right) \quad (1.14)$$

Using the value of  $n$  from equation (1.9) this simplifies to

$$\delta t = \frac{n_0 h_0}{c} \left[ \exp \left( \frac{-h_1}{h_0} \right) - \exp \left( \frac{-h_2}{h_0} \right) \right] \quad (1.15)$$

Taking a typical example where  $h_1=6$  kms and  $h_2=10$  kms ( as in the case of an actual EAS), the duration of the light flash turns out to be  $\sim 1.3$  ns. However, this value is found to be  $\sim 5$  ns if the development of an EAS shower is simulated through Monte Carlo methods. The discrepancy between these two values is expected because of the fact that an oversimplified situation has been used to derive the above equations.

### 1.3.3 Extinction of Cherenkov light in the atmosphere

On their way to the ground through several kilometers, the Cherenkov photons, undergo scattering and absorption processes leading to an exponential extinction. The main dominant sources of extinction are :

#### **Rayleigh Scattering**

Scattering of photons by particles which are of smaller wavelength is called the Rayleigh scattering. The extinction coefficient  $\alpha_R$  is inversely proportional to fourth power of wavelength ( $\alpha_R \propto \lambda^{-4}$ ). Thus mostly the photons with small wavelengths are scattered.

#### **Mie Scattering**

Photons are scattered by the target particles like aerosols with dimensions similar to wavelength. The dependency of the extinction coefficient on the wavelength is ( $\alpha_M \propto \lambda^{-1 \dots 1.5}$ ).

#### **Ozone Absorption**

At wavelengths below 300 nm, ozone is a significant absorber. The fission of ozone ( $O_3 + \gamma \rightarrow O_2 + O$ ) absorbs nearly all photons with wavelengths 200-300nm.

The total attenuation as a result of the above processes is demonstrated pictorially in Fig 1.6. The curves I, II and III represent the total photons available at  $\sim 10$

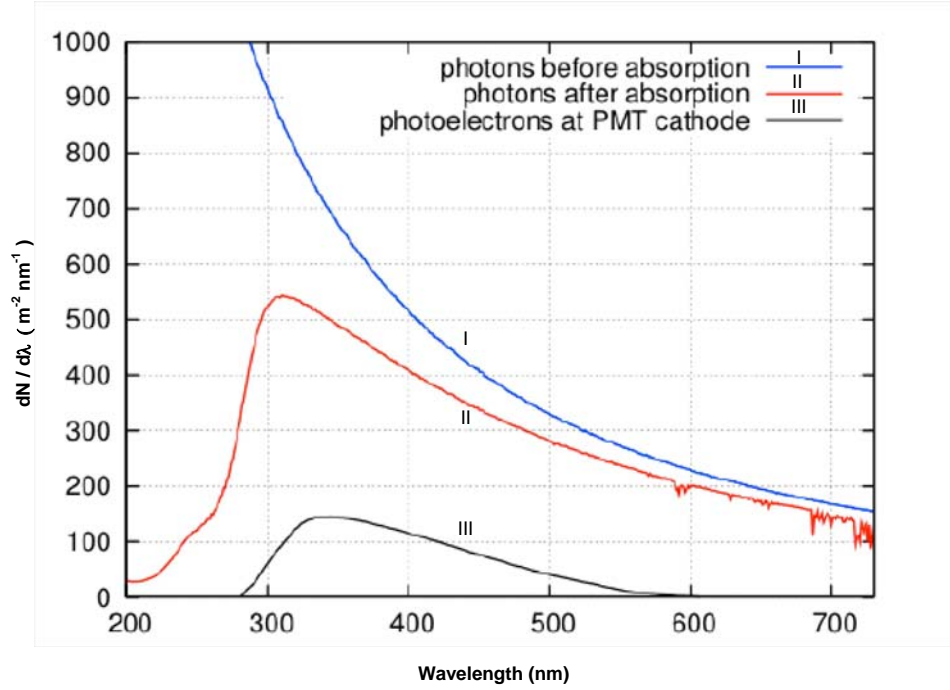


Figure 1.6: Cherenkov photon spectrum before and after attenuation. Adapted from R.M.Wagner Ph.D thesis Munchen University, 2006.

km altitude, photons available at the ground level after absorption and eventually at the PMT cathode, respectively.

## 1.4 Atmospheric imaging Cherenkov technique

### 1.4.1 Detection principle

The technique essentially employs a wide field optical telescope consisting of a large reflector with fast photomultiplier tubes (PMTs) based imaging camera in its focal plane. The light collector of the telescope, located within the Cherenkov light pool can focus light onto a camera in its focal plane and thereby detect the shower. For this purpose the telescope's mirror area needs to be large enough in order to detect a sufficient number of photons from the shower to discriminate against optical background light. The effective detection area of a Cherenkov telescope is given by the area of the Cherenkov light pool itself which is typically of the order of  $10^4 m^2$ .

The camera at the focal plane, records the Cherenkov light photon distribution in the form of images. The pre-requisite for this is to have a fine grained camera

which can resolve details of the shower development to be able to reconstruct shower characteristics and to uncover the differences between the hadronic and electromagnetic showers. For this reason the cameras for Cherenkov telescopes consist of PMTs arranged in a matrix form to collect the reflected light. Furthermore, one needs to have a large field of view (FOV) to allow for observations of extended sources, and a very short light integration time in order to be sensitive to short duration pulses of light produced ( $\sim 5 \text{ ns}$ ). Apart from this, some online trigger schemes are also employed which allow to identify the brief and compact Cherenkov images and to reject light of night sky (LONS) background. The first significant application of

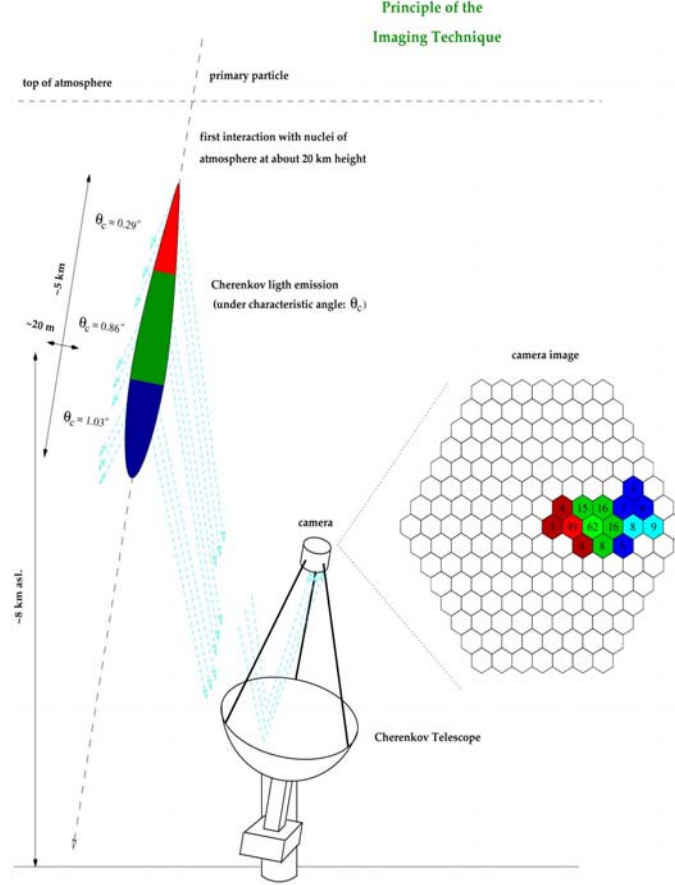


Figure 1.7: Detection of  $\gamma$ -rays through the atmospheric Cherenkov technique. Adapted from K.Daniel Ph.D Thesis Munchen University, 2002.

the atmospheric Cherenkov technique to  $\gamma$ -ray observations was carried out by the Crimea group[1.13]. The principle on which this detection technique is based is illustrated in Fig 1.7. The telescope gets triggered when Cherenkov photons produced in a  $\gamma$ -ray / hadron induced EAS (signal) arrive at the telescope amidst noise of the night-sky photons (background). As discussed earlier, the Cherenkov emission angle is  $\sim 1.36^\circ$  at sea level and decreases with altitude with the result that



most of the Cherenkov photons hit the ground within a circle of radius  $\sim 150\text{m}$ .

With the expectation that the imaging technique would provide vital clues about the identity of primary particle, imaging technique was originally used for improving the angular resolution of the atmospheric Cherenkov telescopes. The idea of the imaging Cherenkov telescopes was proposed in the late 1970s by the Whipple group [1.14]. The Whipple 10m diameter telescope employed the first imaging system and operated for more than 10 years with a camera of 37 photomultiplier tubes. The detection of a  $\gamma$ -ray signal from the Crab Nebula by the Whipple system [1.15] showed for the first time that the imaging Cherenkov telescope does reject the cosmic-ray background and can have an angular resolution of as good as  $\sim 0.1^\circ$

The image of the light measured with the telescope basically represents the ‘shape’ of the electron-positron cascade when it is at its maximum development stage. The shape is elongated along the shower axis, parallel to the direction of the source at which the telescope is pointed. Apart from developing earlier in the atmosphere, Cherenkov photons produced in a  $\gamma$ -ray cascade also have more regular extent (at the detector level). In addition they do not have penetrating particles, like muons, which tend to make the image of a hadron broader and more irregular. The basic differences in the appearance of the Cherenkov light images of a  $\gamma$ -ray shower (centered in the field of view) and that of a typical background shower coming from a random direction within the telescopes field of view are the following : The smaller transverse momenta in electromagnetic interactions, compared to those in hadronic interactions, implies that the secondary particles in the  $\gamma$ -ray shower are, on an average, closer to the direction of the primary. Also, since there is no penetrating component, the local contribution to the light is small and fluctuations in the shower images are far less. The orientation of the roughly elliptical image of a shower depends on the angle which the shower axis makes with the optical axis of the telescope. A shower which has its axis parallel to the optical axis of the telescope and lies upto 150m away from it forms an elliptical image with its major axis pointing towards the camera centre in the focal plane of the telescope. This holds true for both  $\gamma$ -rays and hadrons, but it is easier to characterize the axis of the  $\gamma$ -ray shower because of the fact that it is a more compact image. A schematic comparison of the  $\gamma$ -ray and hadron initiated Cherenkov images is shown in Fig 1.8. For each image, which is essentially elliptical in shape, several image parameters can be evaluated to characterize its shape and orientation. Detailed investigation of the intrinsic differences between electromagnetic and hadronic cascade development by Hillas [1.16], in terms of both longitudinal and transverse development, is crucial to the design of strategies which optimize separation of these two classes. In practice, the most predominant source of events detected by Cherenkov telescopes arises from the proton primaries.

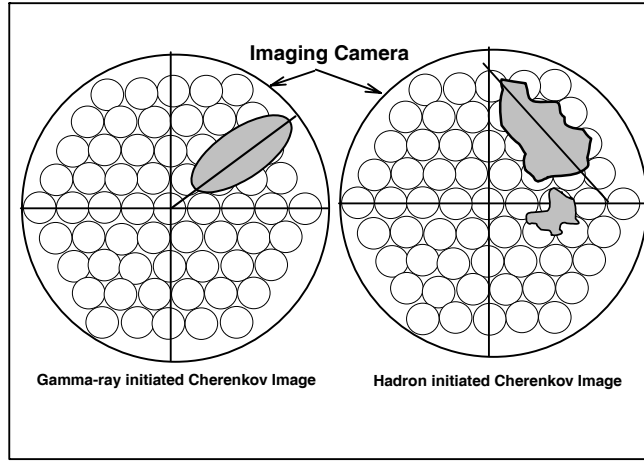


Figure 1.8: Comparison of Cherenkov images initiated by a  $\gamma$ -ray and a hadron initiated shower, in the focal plane of the telescope.  $\gamma$ -ray images are roughly elliptic with major axis pointing towards center of camera, while as hadron images are irregular with major axis oriented randomly.

### 1.4.2 Cherenkov image parameters

A Cherenkov imaging telescope records the arrival direction of the individual Cherenkov photons and the appearance of the recorded image depends upon a number of factors like the nature and the energy of the incident particle, the arrival direction and the impact point of the particle trajectory on the ground. The principle of detecting  $\gamma$ -rays through the imaging technique is depicted further in Fig 1.9a and Fig 1.9b. Segregating the very high-energy  $\gamma$ -ray events from their cosmic-ray counterpart is achieved by exploiting the subtle differences that exist in the two dimensional Cherenkov image characteristics (shape, size and orientation) of the two event species. Gamma-ray events give rise to shower images which are preferentially oriented towards the source position in the image plane. Apart from being narrow and compact in shape, these images have a cometary shape with their light distribution skewed towards their source position in the image plane and become more elongated as the impact parameter increases. On the other hand, hadronic events give rise to images that are, on average, broader and longer and are randomly oriented within the field of view of the camera. For each image, which is essentially elliptical in shape, Hillas parameters [1.17] are calculated to characterize its shape and orientation. The parameters, as depicted in Fig 1.9c, are obtained using moment analysis and are defined as :

**LENGTH**– The rms spread of light along the major axis of the image (a measure of the vertical development of the shower).

**WIDTH** – The rms spread of light along the minor axis of the image (a measure of

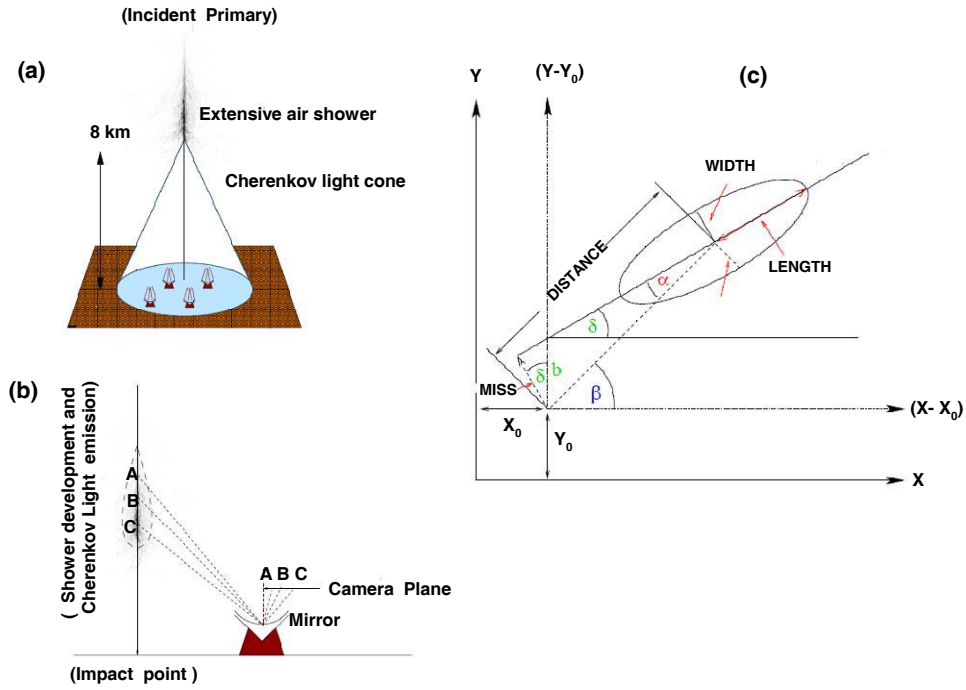


Figure 1.9: (a) The principle of detecting  $\gamma$ -rays through the imaging technique (b) Formation of Cherenkov image in the focal plane. (c) Definition of Hillas parameters characterizing each image and used for rejecting the cosmic-ray background. The ellipse represents the approximate outline of the shower image in the focal plane of the telescope.

the lateral development of the shower).

**DISTANCE**– The distance from the centroid of the image to the centre of the field of view.

**( $\alpha$ )**–The angle between the major axis of the image and a line joining the centroid of the image to the position of the source in the focal plane.

**SIZE** – Sum of all the signals recorded in the clean Cherenkov image.

**FRAC2**– The degree of light concentration as determined from the ratio of the two largest PMT signals to sum of all signals (also referred to as Conc).

In the pioneering work of the Whipple Observatory only one parameter (AZWIDTH) was used in selecting  $\gamma$ -ray events. Later, the technique was refined to Supercuts / Dynamic Supercuts procedure where cuts based on the WIDTH and LENGTH of the image as well as its orientation are used for segregating the  $\gamma$ -rays from the background cosmic-rays.

### 1.4.3 Energy threshold and flux sensitivity

Energy threshold ( $E_{th}$ ) is an important specification of a Cherenkov telescope and is defined as the minimum  $\gamma$ -ray energy for which the signal-to-noise ratio is sufficient to adequately trigger the instrument. Expression for  $E_{th}$  can be determined by evaluating the number of photoelectrons ( $S_{pe}$ ) produced in the light-sensitive PMT due to the Cherenkov signal and the corresponding noise level ( $\sqrt{N_{pe}}$ ) which is due to the fluctuations in Light of night sky (LONS) background induced shot-noise. Using the above definition it can be found that

$$E_{th} \approx \frac{1}{y_\gamma} \left( \sqrt{\frac{\phi_{LONS} \times \Omega \times \tau}{A_c \eta_{pmt}}} \right) \quad (1.16)$$

where  $y_\gamma = \rho_\gamma/E = 0.065$  photons  $m^{-2}GeV^{-1}$  (in the wavelength range 300 to 700 nm) is a scaling factor for expressing Cherenkov photon density ( $\rho_\gamma$ ) in terms of the energy  $E$  of the primary. The typical values for the LONS flux being  $\phi_{LONS} \sim (2 \times 10^{12} \text{ photons } m^{-2}s^{-1}sr^{-1})$  photons in the above specified wavelength range;  $\tau$  is the integration time;  $A_c$  is the collection area of the light-collector;  $\eta_{pmt}$  is the quantum efficiency of the PMT and  $\Omega$  is the field of view of the telescope. In the simple formalism, we ignore the wavelength dependence of the Cherenkov and LONS photons. In reality, however, both  $S_{pe}$  and  $N_{pe}$  are determined by an integral over terms with explicit wavelength dependence and convoluted with the spectral response of the detector. From the above equation, it is clear that we can lower the energy threshold of a Cherenkov telescope by reducing the noise contribution from LONS (i.e. choosing minimum possible  $\Omega \times \tau$ ). Since the Cherenkov light in an air shower has a finite temporal and angular spread, there are limitations to decreasing the energy threshold by reducing the field of view or the integration time. The most straightforward method to reduce the threshold is by using larger mirror area and/or using photodetectors with higher quantum efficiency.

Another important parameter of a Cherenkov telescope is its  $\gamma$ -ray flux sensitivity. In the limit of no cosmic-ray background, the flux sensitivity would be simply determined by the collection area of the telescope as a function of the energy and the observation time. However, since the source has to be detected in the presence of a large isotropic cosmic-ray background, the sensitivity of the instrument is determined by its ability to detect a  $\gamma$ -ray signal in the presence of this cosmic-ray background.

Let us assume that the telescope has the power to discriminate between  $\gamma$ -ray and hadron-generated Cherenkov events. Defining  $\eta_\gamma$  and  $\eta_p$  as the  $\gamma$ -ray and background cosmic-ray retention factors, after the events recorded by the telescope have been selected on the basis of some particular feature of the recorded event. The ‘figure of merit’ or the ‘Quality factor’ QF of a telescope is defined as:

$$QF = \frac{\eta_\gamma}{\sqrt{\eta_p}} \quad (1.17)$$

For generation-I systems, since there is no way to discriminate between  $\gamma$ -ray and background cosmic-ray generated Cherenkov events, we have  $\eta_\gamma=1$ ,  $\eta_p=1$  and hence  $QF=1$ . Ideally one would have liked  $\eta_\gamma \rightarrow 1$ ,  $\eta_p \rightarrow 0$ , so that  $QF \gg 1$ . In other words, it means almost all the cosmic-ray background events are rejected while most of the  $\gamma$ -ray generated events are retained. Generation-II systems and their more recent modifications attempt to achieve this discrimination. Defining  $N_\sigma$  as the minimum number of standard deviations at which a source must be detected to be acceptable, the expression for  $N_\sigma$  is written as:

$$N_\sigma \propto QT^{0.5} E^{(\alpha_p/2 - \alpha_\gamma)} \frac{A_\gamma(E)}{\sqrt{A_p(E)\Omega}} \quad (1.18)$$

where  $A_\gamma(E)$  and  $A_p(E)$  are the collection areas for  $\gamma$ -rays and background cosmic-rays respectively;  $\Omega$  is the solid angle of the telescope;  $T$  is the observation time; and  $\alpha_\gamma$  and  $\alpha_p$  ( $\sim 1.7$ ) are the integral spectral indices for the  $\gamma$ -ray and background cosmic-ray fluxes, respectively. Analysis of the above equation shows that there are several ways to optimize the flux sensitivity of a  $\gamma$ -ray telescope ( i.e, to maximize  $N_\sigma$ ). (a) For typical celestial  $\gamma$ -ray sources where  $\alpha_\gamma > 0.85$ , improvement in sensitivity can be achieved by operating the telescope at the lowest possible threshold energy with the bonus that it allows the detection of  $\gamma$ -ray sources which may have intrinsic energy cutoffs. For  $\gamma$ -ray sources with extremely hard spectra ( $\alpha_\gamma < 0.85$ ), we can improve the sensitivity by operating the telescope at higher energies. (b) Significance of a source detection can also be improved by increasing the observation time. This improvement scales as  $T^{1/2}$  and is applicable for steady sources. (c) Significance also can be substantially improved by using some technique which could reject Cherenkov events due to background cosmic-rays i.e. accomplish  $QF \gg 1$ . This is the backbone of generation-II systems, where  $QF > 7$  can be achieved by applying appropriate background rejection methods.

## 1.5 Challenges of $\gamma$ -ray astronomy

The development of the field of TeV  $\gamma$ -ray astronomy has been an uphill task. Indirect nature of the detection technique along with a need to discriminate the  $\gamma$ -ray signal against a huge background of cosmic-ray events makes it a difficult task. In ground based  $\gamma$ -ray astronomy the primary photon is detected indirectly. The  $\gamma$ -ray initiates an electromagnetic cascade through interaction high in the Earth's atmosphere which acts as a giant local calorimeter. For primary energies  $\sim < 20$  TeV the cascade dies out in the upper atmosphere but the Cherenkov light penetrates to ground level where it may be collected as an image with a suitable arrangement of focussing mirrors. Although subject to fluctuations at the point of origin of the cascade and its subsequent development, the Cherenkov light carries with it the information pertaining to the exact direction of origin of the primary  $\gamma$ -ray on the

celestial sphere along with the energy of the initiating primary. Calorimetry should therefore be possible if the Cherenkov light can be detected in some unambiguous manner. Unfortunately, since the  $\gamma$ -ray signal has to be picked out against a generally overwhelming background of hadronic cascades induced by the interactions of protons and the stripped atomic nuclei. It has long been realized that if TeV  $\gamma$ -ray astronomy is to be viable and a catalogue of statistically significant cosmic point sources of TeV photons established, then it has to be capable of rejecting this unwanted hadronic background at a remarkably high level of efficiency, while simultaneously retaining a majority of the genuine  $\gamma$ -ray events. The success of TeV  $\gamma$ -ray astronomy, at energies  $> 100\text{GeV}$ , thus mainly depends upon how well the  $\gamma$ -ray and hadronic cascades can be separated based on the atmospheric Cherenkov and/or any other promising techniques.

The primary objective of  $\gamma$ -ray astronomy in the energy range  $100\text{GeV}$ - $10\text{TeV}$  therefore mainly becomes a task of detecting  $\gamma$ -rays in the presence of a dominant background of hadronic showers, which must be rejected with very high efficiency if the signal-to-noise ratio of the detection process is to be improved. Extrapolation of known cosmic point source fluxes of  $\gamma$ -rays at several GeV energies into the TeV range indicates the futility of attempting detection based solely on unprocessed data. Strategies developed to date, with capability of greater than 99% rejection of background, exploit three distinct aspects of the ACT process, namely: (a) developmental and evolutionary differences of electromagnetic and hadronic cascades in the atmosphere, (b) consequential differences in the process of mapping Cherenkov photons at ground level, by suitable focusing optics, and (c) registration of the mapped photons or formation of a picture or image, through employment of some appropriate, sensitive, fast, wide-angle focal plane detector.

While some minimal discrimination has been shown to be possible in hardware at the detection stage, the predominantly successful strategies which have so far evolved have been based on selection and analysis of  $\gamma$ -ray candidate events subsequent to detection.

## 1.6 Milestones and present status of the field

The field attained a firm experimental footing with the development of IACT, which provides a method to effectively discriminate between the  $\gamma$ -ray initiated showers and the background cosmic-ray showers. A number of IACTs have been put around the world, which includes Durham, CANGAROO, Telescope Array, Crimean Astrophysical Observatory, SHALON, TACTIC, Whipple, HEGRA and CAT. The 1990's saw two particularly important developments: the detection of the first extragalactic sources by the Whipple Collaboration, starting with the blazars Mrk 421 [1.18] and Markarian 501 [1.19], and the application of the first stereo imaging technique by the HEGRA array [1.20]. Despite this progress, the relative scarcity of bright TeV  $\gamma$ -ray sources ( $< 10$  were identified by 2000) highlighted the necessity

for improved instrumentation. Cherenkov wavefront samplers such as CELESTE and STACEE attempted to probe to lower energies, and hence higher  $\gamma$ -ray fluxes and larger distances, using converted solar farms; however, the difficulty of discriminating  $\gamma$ -rays from the cosmic-ray background using this technique limited its effectiveness. Starting with the commissioning of H.E.S.S. in 2003, the new generation of IACTs - H.E.S.S., MAGIC and VERITAS - have provided the required order of magnitude improvement in sensitivity, and firmly established  $\gamma$ -ray studies as an important astronomical discipline.

Presently three major stereoscopic imaging atmospheric Cherenkov telescope systems are in operation globally. H.E.S.S., located in the Khomas Highland of Namibia ( $-23^{\circ}\text{N}$ ,  $-16^{\circ}\text{W}$ , altitude 1800 m), consists of four telescopes arranged on a square with 120 m side length. Each telescope has a mirror area of  $107\text{ m}^2$  and is equipped with a 960 pixel camera covering a  $\sim 5^{\circ}$  FOV. Recently a 28 m diameter IACT with a 2048 pixel imaging camera has been installed at the center of the array. VERITAS, at the Fred Lawrence Whipple Observatory in southern Arizona ( $32^{\circ}\text{N}$ ,  $111^{\circ}\text{W}$ , altitude 1275 m) has similar characteristics, with 4 telescopes of  $107\text{ m}^2$  area and 499-pixel camera, covering  $\sim 3.5^{\circ}$ . MAGIC ( $28^{\circ}\text{N}$ ,  $17^{\circ}\text{W}$ , altitude 2225 m) originally consisted of a single, very large reflector ( $236\text{ m}^2$ ) on the Canary island of La Palma, with a  $\sim 3.5^{\circ}$  camera. In 2009, a second telescope with the same mirror area was installed at a distance of  $\sim 85\text{ m}$  from the first. The catalog of TeV sources has grown rapidly with the commissioning of H.E.S.S. It has continued to expand in recent years with MAGIC and VERITAS becoming operational and the number of sources has now reached  $> 160$  sources. The source catalog can broadly be categorized into two classes : The galactic and the extragalactic sources.

### 1.6.1 Galactic sources

There are presently  $\sim 110$  known TeV  $\gamma$ -ray sources within our Galaxy, which are indicated by their association with known Galactic sources at other wavelengths [1.21]. Galactic plane is densely populated with TeV  $\gamma$ -ray sources, primarily clustered within the inner  $\pm 60^{\circ}$ , in the galactic longitude (Fig 1.10). Many Galactic  $\gamma$ -ray sources are extended, allowing detailed studies of source morphology and spatially resolved spectra, while others are time variable and/or periodic. The main sub source classification can be (a) Galactic center sources (b) Supernova Remnants (SNR) (c) Pulsar and pulsar wind Nebulae (PWN) (d) Compact object binary systems (e) stellar clusters and (f) Unidentified VHE  $\gamma$ -ray sources. A brief discussion on these sub-classes is presented below.

In the galactic center a TeV  $\gamma$ -ray emitter has been reported by many imaging telescopes. Determining the nature of this source is a difficult task, due to the complexity of the region, which includes multiple potential counterparts. An additional, faint, component is also seen, which extends in both directions along the galactic plane. The extended component is spatially correlated with a complex

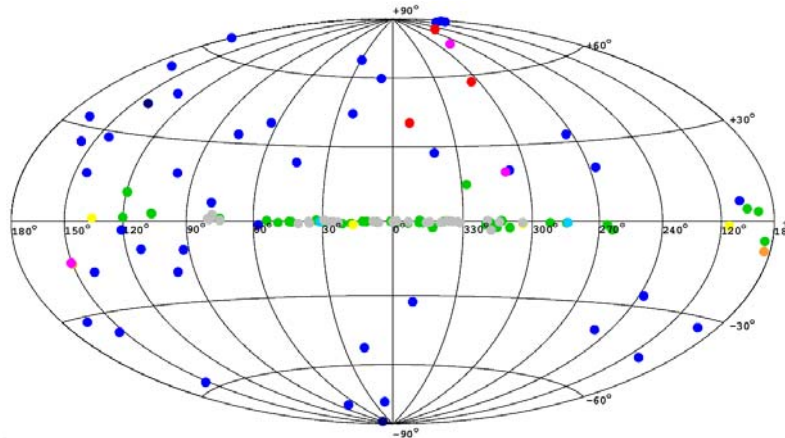


Figure 1.10: Catalogue of VHE sources observed by ground-based Cherenkov telescopes. The TeGeVcat is collecting all the information publicly available about TeV sources observed by the past generation and current generation of imaging Cherenkov telescopes. Adapted from The TeGeV Catalogue ASDC.

of giant molecular clouds in the central 200 pc of the Milky Way, and the TeV emission can be attributed to the decay of neutral pions produced in the interactions of hadronic cosmic-rays with the cloud material. The Galactic Centre is also a prime candidate region to search for  $\gamma$ -ray emission due to dark matter particle self-annihilation. The Crab Nebula is the nearby remnant of a historical supernova explosion at a distance of  $\sim 2$  kpc, observed in 1054 A.D. There is no detected shell, and the broadband emission below  $\sim 100$  MeV is dominated by a bright synchrotron nebula, powered by a central pulsar (PSR B0531+21). This pulsar is the most energetic pulsar in our Galaxy, with a pulse period of 33 ms, and a spin-down power of  $4.6 \times 10^{38} \text{ erg s}^{-1}$ . The Crab Nebula and Pulsar hold a unique place in the development of TeV astronomy: The birth of the field as an astronomical discipline, can be traced to the detection of the Crab Nebula TeV source by the Whipple 10 m telescope, in the first application of the imaging atmospheric Cherenkov technique. Subsequently, the Crab Nebula has acted as a bright, standard candle for TeV  $\gamma$ -ray observatories. However, recently multiple day-scale flaring events at energies below 1 GeV have been detected by Large Area Fermi Gamma-ray space telescope [1.22-1.23]. Measurements from Cherenkov imaging telescopes do not support these results, but are not necessarily in conflict, because of differing duty cycles.

Pulsar wind nebulae (PWN) are the most abundant class of known VHE emitters in the Galaxy, with  $\sim 35$  firm examples, and numerous other sources where



the PWN association is more tentative (for reviews see e.g. [1.24-1.28]). The essential emission mechanisms i.e, shock accelerated leptons producing synchrotron and inverse Compton radiation have already been described for the case of the Crab PWN, but the Crab Nebula is far from the typical object. Understanding of the structure and evolution of PWN has advanced significantly over the past few years, in particular thanks to the high resolution X-ray imaging provided by Chandra X-ray telescope. Supernova remnants, are believed to efficiently accelerate particles at the shock front where the expanding SNR encounters the surrounding medium (e.g. [1.29]). This likely occurs through diffusive shock acceleration (first order Fermi acceleration), in which charged particles are reflected from magnetic inhomogeneities and repeatedly cross the shock front, gaining energy with each crossing.

The process of diffusive shock acceleration is not limited to supernova remnant shells. An alternative scenario invokes particle acceleration at the shock formed by the collision between the supersonic stellar winds of massive stars in close binary systems. Stellar winds may also become collectively important in large assemblies of massive stars. The combined effect of the stellar winds, coupled with the effect of multiple SNRs, results in an overall wind from the cluster which forms a giant superbubble in the interstellar medium. Particle acceleration can occur where the cluster wind interacts with the surrounding medium. Massive star associations are naturally likely to host other potential source counterparts for TeV emission, such as compact object binary systems, individual supernova remnants and pulsar wind nebulae. Despite many early unconfirmed claims, the first definitive detection of a TeV  $\gamma$ -ray binary system was not published until 2005 [1.30]. The population has grown slowly, and now consists of four clearly identified systems, plus marginal evidence for transient emission associated with Cyg X-1. The  $\gamma$ -ray emission from binaries is believed to be powered either by accretion (most likely onto a black hole), or by a pulsar wind.

### 1.6.2 Extragalactic sources

Approximately 1% of all galaxies host an active nucleus; a central compact region with much higher than normal luminosity. Around 10% of these Active Galactic Nuclei (AGN) exhibit relativistic jets, powered by accretion onto a supermassive black hole. Many of the observational characteristics of AGN can be attributed to the geometry of the system; in particular, the orientation of the jets with respect to the observer. Blazars, which host a jet oriented at an acute angle to the line of sight, are of particular interest for  $\gamma$ -ray astronomy, as the emission from these objects is dominated by relativistic beaming effects, which dramatically boost the observed photon energies and luminosity. The first extragalactic source discovered at TeV energies was Markarian 421, a blazar of the BL Lacertae sub-class. The extragalactic TeV catalog now comprises  $\sim 50$  objects, and continues to increase steadily. The majority ( $\sim 80\%$ ) of the known TeV blazars are high-frequency peaked

objects (HBL). The mechanisms which drive the high energy emission from blazars remain poorly understood. Briefly, in leptonic scenarios, a population of electrons is accelerated to TeV energies, typically through Fermi acceleration by shocks in the AGN jet. These electrons then cool by radiating X-ray synchrotron photons. TeV emission results from inverse Compton interactions of the electrons with either their self-generated synchrotron photons, or an external photon field. The strong correlation between X-ray and TeV emission which is often observed provides evidence for a common origin. Hadronic models are less favoured, because the cooling times for the relevant processes are long, making rapid variability difficult to explain. Many of the AGN detected at TeV energies exhibit extreme variability. The timescales can range from minutes to days to months, and the observed flux can change by more than an order of magnitude. Such rapid variability can be used to place constraints on the size of the emission region. Most distant object detected is 3C279 [1.31], with a relatively modest redshift of  $z \sim 0.5362$ . The population is truncated at large distances due to the absorption of TeV  $\gamma$ -rays by electron-positron pair production with the low energy photons of the extragalactic background light (EBL). This effect is energy dependent, and can thus strongly modify the observed VHE spectra of extragalactic sources. While this limits the observation of distant TeV sources, it provides a mechanism by which to infer the intensity of the EBL, by applying assumptions about the intrinsic spectra at the source [1.32]. TeV blazar observations have also been suggested as probes of other physical phenomena, such as the acceleration and propagation of ultra-high energy cosmic rays etc [1.33-1.34]. Nearby radio galaxies are another class of extra-galactic TeV  $\gamma$ -ray sources. These are the sources in which the jet is not directly oriented towards the line-of-sight, but provide an alternative method by which to investigate the particle acceleration and  $\gamma$ -ray emission from relativistic outflows in AGN. The advantage of studying such objects lies in the fact that these jets can be resolved from radio to X-ray wavelengths, allowing the possibility of correlating the  $\gamma$ -ray emission state with observed changes in the jet structure. Three radio galaxies have been identified as TeV emitters: M 87, Centaurus A and NGC 1275. M87 is the most well studied of these, and was first reported as a  $\gamma$ -ray source by the HEGRA collaboration [1.35]. M 87 is a giant radio galaxy at a distance of  $\sim 16.7$  Mpc, displaying a prominent misaligned jet, with an orientation angle of  $\leq 20^\circ$  to the line-of-sight. Starburst galaxies are yet another important class of extragalactic sources. These exhibit an extremely high rate of star formation, sometimes triggered by interaction with another galaxy. High cosmic-ray and gas densities in the starburst region make these objects promising targets for  $\gamma$ -ray observations, with emission predicted to result from the interactions of hadronic cosmic-rays in the dense gas. TeV emission has now been identified from two starburst galaxies: M 82 [1.36] and NGC 253 [1.36]. M 82 is a bright galaxy located at a distance of  $\sim 3.9$  Mpc, with an active starburst region at its centre. The star formation rate in this region is approximately 10 times that of the Milky Way, with an estimated supernova rate of  $\sim 0.1$  to  $0.3$  per

year. In addition to above, there are some other important classes of sources which deserve a mention e.g., the Large Magellenic Cloud which however are predicted to give fluxes below the current instrument sensitivity in the TeV domain.

A summary of the main galactic and extragalactic sources is presented in Table 1.3 below

Source type	Total detected	Source type	Total detected
<i>Binary</i>	5	<i>Shell</i>	13
<i>SNR</i>	9	<i>FSRQ</i>	3
<i>HBL</i>	41	<i>Massive star cluster</i>	4
<i>LBL</i>	1	<i>XRB</i>	3
<i>IBL</i>	7	<i>UID</i>	30
<i>PWN</i>	35	<i>FRI</i>	3
<i>PSR</i>	4	<i>SNR/Molec. cloud</i>	9

Table 1.3: Various types of TeV sources; SNR=Supernova remnant; HBL=High energy peaked BL lac object; LBL=Low energy peaked BL lac object; IBL= Intermediate energy peaked BL lac object; PWN=Pulsar wind nebulae; PSR= Pulsar; FSRQ=Flat spectrum radio quasar; XRB=X-ray binary; UID=Unidentified sources; FRI= Fraunhofer radio galaxies

## 1.7 References

- [1.1] E. Rutherford and F. Cook, Physics Review 16 (1903) 183.
- [1.2] A. Gockel et al., Phys Zeits 11 (1910) 280.
- [1.3] R. Gould and G. Schreder, Physics Review 155 (1967) 1408.
- [1.4] F. Steker and O.D. Jager, Journal of Astrophysics 415 (1993) L71.
- [1.5] T.C. Weekes, Physics reports 160 (1988) 1.
- [1.6] H. Bethe and W Hitler, Quantum theory of radiation 3rd ed Dover Press 1954.
- [1.7] D. Paneque. Jour of Phys Conf series 375 (2012) 052020.
- [1.8] A.A. Abdo, et al., Arxiv:1108.1435 (2011).
- [1.9] O. Heaviside, Electrical Papers 2 (1890) 494.
- [1.10] L. Mallet, C R Academy of sciences(Paris) 183 (1926) 274.
- [1.11] P.A. Cherenkov, Physics Review 52 (1937) 378.

- [1.12] N.L. Balazs ,Phys Review 104 (1956) 1220.
- [1.13] A.E.Chudakov, et al., Proc. of P. N. Lebedev Phys. Inst., 26,(1965)118.
- [1.14] G.G. Fazio et al., Ap.J 175 (1972) L117.
- [1.15] T.C. Weekes et al., Ap.J 342 (1989) 379.
- [1.16] A.M. Hillas, Space Science Reviews, 75 (1996) 17.
- [1.17] A.M. Hillas, Proc 19th ICRC, La Jolla 3 (1985) 445.
- [1.18] M. Punch et al., Nature 358 (1992) 477.
- [1.19] J. Quinn et al., ApJL 456 (1996) L83.
- [1.20] A. Daum et al., Astropart. Phys. 8 (1997) 1.
- [1.21] E. de O Wilhelmi et al., American Institute of Physics Conference Series, 1357 (2011) 213.
- [1.22] A.A. Abdo et al., Science 331 (2011) 739.
- [1.23] R. Ojha et al., ATel 4855 (2013).
- [1.24] <http://tevcat.uchicago.edu> .
- [1.25] S. Funk, arXiv:1204.4529v1 (2011).
- [1.26] F.M. Rieger et al., arXiv:1302.5603v1 (2013).
- [1.27] J.A. Hinton et al., arXiv:1006.5210v2 (2010).
- [1.28] T.C. Weekes, arXiv:0811.1197v1 (2008).
- [1.29] L.O. Drury, F. A. Aharonian, H. J. Voelk, A&A 287 (1994) 959.
- [1.30] F. Aharonian, et al., ArXiv:0506.280v2 (2005).
- [1.31] J. Aleksic, et al., A&A 530 (2011) A4.
- [1.32] F.W. Stecker, O. C. de Jager, M. H. Salamon, ApJL 390 (1992) L 49.
- [1.33] F.A. Aharonian, P. S. Coppi, H. J. Voelk, ApJL 423 (1994) L5.
- [1.34] S. Gabici, F.A. Aharonian, Ap & SS 309 (2007) 465.
- [1.35] F. Aharonian et al., A&A 403 (2003) L1.
- [1.36] V.A. Acciari et al., Nature 462 (2009) 770.
- [1.37] F. Acero et al., Science 326 (2009) 1080.

## Chapter 2

# TACTIC telescope : Details and Performance

The TACTIC (TeV Atmospheric Cherenkov Telescope with Imaging Camera)  $\gamma$ -ray telescope [2.1] has been in operation at Mt. Abu ( $24.6^\circ$  N,  $72.7^\circ$  E, 1300m asl), India, for the last several years. The telescope uses a tessellated light-collector of area  $\sim 9.5\text{m}^2$  which is capable of tracking a celestial source across the sky. It deploys a 349-pixel imaging camera, with a uniform pixel resolution of  $\sim 0.3^\circ$  and a  $\sim 6^\circ \times 6^\circ$  field-of-view, to take a fast snapshot of the atmospheric Cherenkov events produced by an incoming cosmic-ray particle or a  $\gamma$ -ray photon with an energy above  $\sim 1\text{TeV}$ . The photographs of the TACTIC imaging telescope and its back-end signal processing electronics are shown in Fig 2.1

## 2.1 Observatory site, mechanical assembly and drive control

Keeping in mind the overall scientific requirements of a good astronomical site for atmospheric Cherenkov systems, a comprehensive site survey program was undertaken by us in 1993 [2.2] and Mt.Abu ( $24.63^\circ\text{N}$ ,  $72.75^\circ\text{E}$ , 1300m asl), a hill resort in Western Rajasthan (India) was found to be a suitable location in the country for setting up the TACTIC telescope.

The main sub-assemblies of the TACTIC telescope are : mirror basket, mirror fixing/adjusting frames, zenithal and azimuthal gear assembly, encoders/motors for its two axes, camera support boom assembly and the photomultiplier tube-based imaging camera. A three dimensional truss type structure has been used to support the mirror frame. The mirror basket consists of a 3 layer welded mildsteel tubular grid structure. The individual mirror facets (34 in number, each with a diameter 60 cm and weighing around 20 kg) are supported on three levelling studs, so that desired inclination of the specific mirror with respect to the telescope axis can be

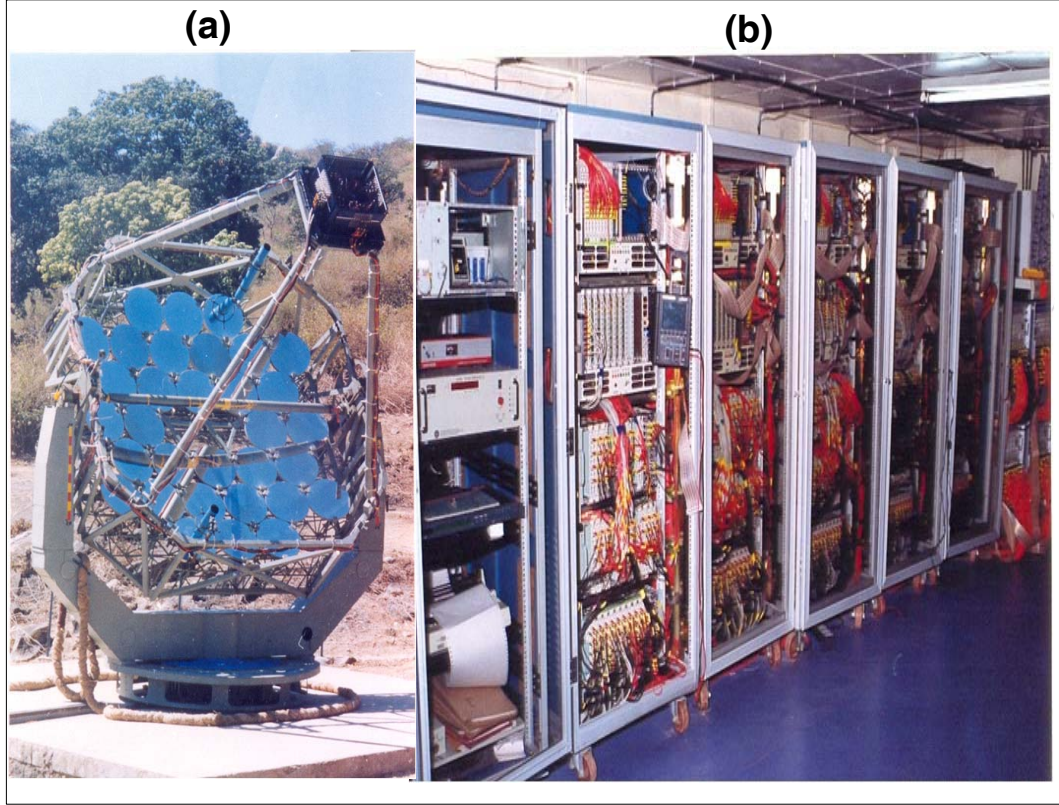


Figure 2.1: (a) Photograph of the 349-pixel TACTIC imaging telescope (b) Photograph of back-end signal processing electronics used in the telescope.

achieved. The weight of the moving part of the telescope is around 6.5 tons. The zenithal motion to the telescope is given from only one end of the basket. Five stage gear box has been used in the zenithal drive of the telescope and a four stage gear box has been used in the azimuth drive. A large capacity circular cable drag chain has been provided for easy and free movement of the nearly 700 signal and high voltage cables.

The need for using a large light collector aperture ( $\sim 3.5\text{m}$ ) and the large telescope weight ( $\sim 6.5$  tons) has led to the choice of an altitude-azimuth (alt-azm) mounting for the TACTIC telescope, as against the comparatively simpler equatorial mounting. The main advantage of the alt-azm mount is that the telescope weight is supported uniformly on a horizontally-placed central thrust-bearing. The telescope uses two 100 N cm hybrid stepper motors (Pacific Scientific make, Model H 31NREB, NEMA Size 34) for driving its azimuthal and zenithal axes through multistage gear-trains. The speed and direction control for each axis of the telescope is implemented through an in-house developed CAMAC-based Stepper Motor Controller. The two resolvers, coupled to 16-bit resolution decoders, were calibrated

on a precision indexing table (resolution  $\sim 0.5$  arc-sec; accuracy  $\sim 5$  arc-sec) and it was found that the error profiles are largely of systematic nature [2.3]. Accordingly, a software-based procedure has been successfully developed for compensating for this systematic error-profile of the encoders. An Artificial Neural Network based error compensation methodology has also been established for this error compensation [2.4]. A GPS based CAMAC-compatible digital clock (Hytec Electronics Ltd. make; GPS92) with a resolution of  $\sim 10$  ns and absolute time accuracy of  $\sim 100$  ns, is used to compute the source co-ordinates in real time. The new co-ordinates of the source are calculated after every second while tracking a candidate source. More details regarding various hardware components of the telescope drive system are discussed in [2.5].

The user friendly in-house developed tracking system software provides an independent movement for the zenithal and azimuthal axes so that a matching between the telescope pointing direction and the source direction is obtained with an accuracy better than  $\pm 2$  arc-min. Once the error goes outside the permissible bounds of  $\leq 2$  arc-min in case of either axis, at a zenith angle  $\geq 7^\circ$ , a correction cycle (in the form of temporary halt or faster movement at a stepping rate of  $\sim 100$  Hz) is applied till the corresponding offset gets restored to within  $\leq 1.0$  arc-min. While this on-off type of correction cycle works perfectly for correcting the zenithal error, irrespective of the source declination, following the same principle in the azimuth axis works properly for only those sources which have a minimum zenith angle of  $7^\circ$ . For sources which have a minimum zenith angle in the range  $2^\circ - 7^\circ$ , we have provided for ramp-up correction cycle (with stepping rates upto 400 Hz), for tracking them close to their upper transit, to avoid the problem of indefinite chase which would have otherwise occurred if the correction was performed at a stepping rate of 100 Hz. Furthermore, since an azimuth error upto  $\sim 15$  arc-min can be easily tolerated at a typical zenith angle of around  $3^\circ$  without leading to any serious deterioration in the pointing direction of the telescope, the permissible azimuth error band has been accordingly dynamically widened from  $\sim 2$  arc-min to  $\sim 15$  arc-min, depending on the zenith angle of the source, so that frequent correction cycles leading to a possible 'hunting' problem can be avoided. Fig 2.2 gives the representative error profiles in the zenith, azimuth and pointing angles of the telescope during tracking of the Crab Nebula (declination  $\sim 22.01^\circ$  N; zenith angle  $\sim 2.60^\circ$  at upper transit) for  $\sim 6$  hours.

The tracking accuracy of the telescope is also checked on a regular basis with so called 'point runs', where a reasonably bright star, having a declination close to that of the candidate  $\gamma$ -ray source is tracked continuously for  $\sim 5$  hours. The point run calibration data (corrected zenith and azimuth angle of the telescope when the star image is centered) are then incorporated in the telescope drive system software so that appropriate corrections can be applied directly in real time while tracking a candidate  $\gamma$ -ray source.

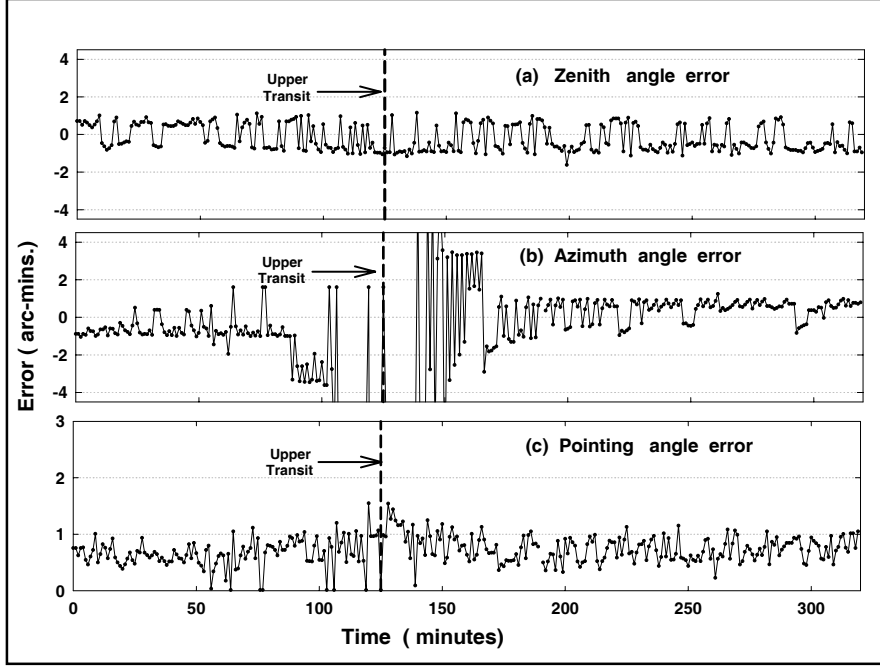


Figure 2.2: Representative error-profiles in the (a) zenith, (b) azimuth and (c) pointing angles of the TACTIC telescope, obtained while tracking the Crab Nebula (declination  $\sim 22.01^\circ$  N; zenith angle  $\sim 2.60^\circ$  at upper transit) for  $\sim 6$ h.

## 2.2 Light-collector design of the TACTIC telescope

The TACTIC light-collector with a collection area of  $\sim 9.5\text{m}^2$  uses 34 front-face aluminium-coated, glass spherical mirrors of 60cm diameter each with the following characteristics (i) focal length  $\sim 400\text{cm}$ , (ii) surface figure  $\sim \text{few } \lambda$  (iii) reflection coefficient  $> 80\%$  at a wavelength of  $\sim 400\text{nm}$  and (iv) thickness 20mm to 40mm. The shorter focal length facets are deployed close to the principal axis of the basket while the longer focal length facets are deployed around the periphery. The peripheral mirrors have the effect of increasing the overall spot size as they function in an off-axis incidence mode.

The alignment of the various mirror facets is done by a two step process. In the first step, the orientation angle and the focal distance of the mirror facet pole is precalculated from geometrical considerations and the orientation of the mirror to within an error of  $\sim 1^\circ$  is adjusted using a dummy facet at all the locations. The orientation of the dummy facet is set by adjusting the 3 ball-joints which couple the triangular mirror holding frame to the mirror basket. In the second step, the individual mirror facets of the telescope are further aligned precisely using an indigenously developed laser plumb-line. With telescope pointed in the vertical



direction, an individual mirror facet is installed at its pre-designated location and the laser plumb line is suspended over it. The mirror facet is then slowly adjusted such that the reflected beam hits the centre of the focal plane. This fine adjustment is done by varying the gaps at the 3 locations below the mirror facet where it rests on the triangular holding frame. The above procedure is repeated for a total of 5 points on each mirror and these preselected points are the pole of the mirror facet and four equidistant points on the periphery of the mirror. Using this technique for all the mirror facets, one at a time, a common focus with the minimum possible image spread was obtained at a focal plane distance of 386cm instead of at 400cm, as would have been expected for a standard paraboloid or a Davies-Cotton design of the reflector. This value of focal plane distance was chosen on the basis of the simulation results [2.6]. The alignment of the mirror facets is further confirmed by observing a bright star image at the focal plane. Gross misalignment in any of the facets is easily identified as it results in multiple images being seen on the focal plane.

In order to evaluate the optical quality of the light collector experimentally, the telescope was pointed towards the bright star  $\zeta$ -Tauri and its image recorded by monitoring the anode current of the central pixel of the imaging camera. The anode current versus angular offset plot is shown in Fig 2.3a. The point-spread function shown has a FWHM of  $\sim 0.185^\circ$  ( $\equiv 12.5\text{mm}$ ) and  $D_{90} \sim 0.34^\circ$  ( $\equiv 22.8\text{mm}$ ). Here,  $D_{90}$  is defined as the diameter of circle, concentric with the centroid of the image, within which 90% of reflected rays lie. An image of the star Sirius recorded at the focal plane of the telescope has also been shown in Fig 2.3b and it has superimposed on it two circles which correspond to the diameter of the pixel and measured  $D_{90}$  value calculated on the basis of Fig 2.3a.

## 2.3 The imaging camera

The camera frame is made up of two 5mm thick aluminium plates in which 19mm diameter holes are drilled at a pitch of 22mm corresponding to the locations of the 349 pixels. The photomultiplier tube is held in place by a metallic collar fixed to its socket which in turn is held to the rear plate by a specially designed fastener. The central pixel of the camera which is on the principal axis of the light collector is used for checking the alignment of the mirror facets and the tracking/pointing accuracy of the telescope. The pixels are numbered sequentially clockwise from the central pixel which is designated as No.1. This arrangement has the advantage of not disturbing the numbering subsequently as the number of operating pixels was increased from 81 in 1997 to 349 in the year 2001. The camera mounting system has a provision of varying its distance from the mirror basket by about  $\pm 20\text{cm}$  which is very useful in optimizing the focal plane distance for obtaining the best possible point spread function.

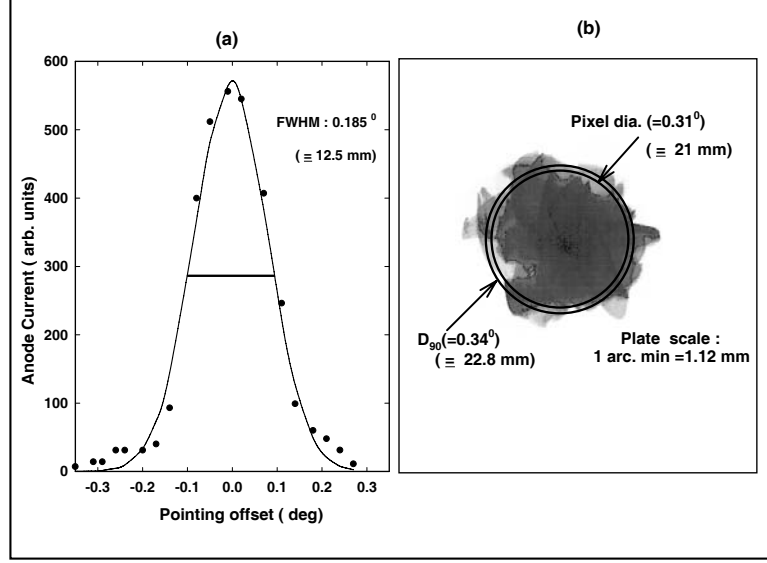


Figure 2.3: (a) Measured point spread function of the TACTIC telescope light collector. (b) Photograph of the image produced by Sirius. Circles superimposed on the image have diameters  $\sim 0.31^\circ$  and  $\sim 0.34^\circ$  and represent the diameter of the camera pixels and  $D_{90}$ , respectively.

### 2.3.1 Photomultiplier tubes and light guides

The imaging camera uses 19mm diameter photomultiplier tubes (ETL-9083 UVB). The bialkali photocathode has a maximum quantum efficiency of  $\sim 27\%$  at 340nm and the use of UV glass for the window has enhanced its sensitivity in the 280-300nm wavelength band. The 10 stage linear focussed photomultiplier tube (PMT) has a rise time of  $\sim 1.8\text{ns}$  which is compatible with the time profile of the Cherenkov pulse. A low current zener diode-based voltage divider network (VDN) is used with the PMT. This VDN design [2.7] has the advantage of ensuring stable voltages at the last two dynodes with VDN current of only about  $240\mu\text{A}$  which is a factor of 5 less than the minimum current recommended for a resistive VDN. The VDN uses negative voltage and the photocathode is at a high voltage of 1000-1400V while the anode is at the ground potential. The VDN of the PMT is permanently soldered to its socket and two RG174 coaxial cables from each VDN circuit board are terminated with coaxial connectors on the connector panels fixed to 4 sides of the camera. The high voltage cable has a plug type SHV connector, while the signal cable uses a BNC connector thereby preventing the possibility of wrong connections.

The Compound Parabolic Concentrator (CPC) shape was chosen for the light guides to ensure better light collection efficiency and reduction in the background light falling on the photomultipliers. After prolonged trials with various materials the light guides were made of SS-304. Some of the important geometrical parameters



interrupts the data acquisition for reading the charge ADC data from all the 349 pixels. The final trigger pulse is also used for latching the GPS-based clock. The scalers and charge-to-digital converters (CDC) for the 349 channels use 5 CAMAC crates each. Each of these crates is controlled using an in-house developed multicrate CAMAC controller and five such controllers are daisy chained and connected to a data acquisition PC. A similar strategy following a custom built standard has been used for the computer-programmable high voltage units.

### 2.4.1 Trigger generation

The imaging camera uses a programmable topological trigger [2.8] which can pick up events with a variety of trigger configurations. As the trigger scheme is not hard wired, a number of coincidence trigger options ( e.g Nearest Neighbour Pairs, Nearest Neighbour Non-collinear Triplets and Nearest Neighbour Non-Collinear Quadruplets) can be generated under software control. The trigger criteria have been implemented by dividing these inner 240 pixels into 20 groups of 12 ( $3 \times 4$ ) pixels. A section of 5 such groups is connected to a TACTIC Trigger Generator (TTG) module and a total of 4 TTGs are required for a maximum of  $15 \times 16$  matrix of trigger pixels. The trigger scheme has been designed around 16k x 4 bit fast static RAM (Toshiba make TC55B417; access time of  $< 8\text{ns}$ ). Each of the 4 TTG modules uses 5 memory ICs and has horizontal cascading built into it. The TTG operation starts with writing of the data, as per a user defined topology, from a disk file into each of its memories under CAMAC control. Apart from generating the prompt trigger, the trigger generator has a provision for producing a chance coincidence output based on  $^{12}C_2$  combinations from various groups of closely spaced 12 channels. This chance coincidence output is used as a system monitor for evaluating its overall functioning during an observation run. Monitoring of the chance coincidence rate has also helped in keeping a close check on the operation of the telescope and the quality of the data collected by it. Other details regarding the design, implementation and performance evaluation of the programmable topological trigger generator for the 349-pixel imaging camera of the TACTIC telescope are discussed in [2.9].

## 2.5 Data acquisition and control system

The data acquisition and control system of the telescope has been designed around a network of PCs running the QNX (version 4.25 [2.10]) real-time operating system. The software is designed for the real time acquisition of event and calibration data and on-line display of telescope status in terms of prompt and chance coincidence rates and the functional status of each of the 349 pixels of the camera. The QNX operating system was chosen for its multitasking, priority-driven scheduling and fast context switching capabilities. In addition, the operating system also provides

a powerful set of interprocess communication capabilities via messages, proxies and signals. The data acquisition and control of the TACTIC is handled by a network of three personal computers. While one PC is used to monitor the scaler rates and control the high voltage (HV) to the photomultipliers, the other PC handles the acquisition of the event and calibration data and the programming of the TTGs. These two front-end PCs, referred to as the rate stabilization node and the data acquisition node respectively, along with a master node form the multinode Data Acquisition and Control network of the TACTIC Imaging telescope. All executable routines stored on the master node are spawned on to the other two front-end nodes as and when required. The same network is extended to two more LINUX-based PCs which are used for on-line data analysis and archiving. An event handler module controls the whole process of data acquisition and also provides the link between the TACTIC hardware and the application software. The event handler accepts the atmospheric Cherenkov events, calibration and chance trigger outputs from various TTG modules and interrupts the front end data acquisition node. The system acquires the 349 channel CDC data for the trigger selected atmospheric Cherenkov events, relative calibration flashes generated by the calibration LED and sky pedestal events, in addition to CDC data for the 4 absolute calibration channels. The high voltage and scaler data are also logged continuously, though at a much lower frequency. At event occurrence the event handler also generates a TTL output for latching the system clock and a 20 ns wide NIM pulse for gating the CDC modules. The dead time of the system has been experimentally measured to be  $\sim 2.5$ ms by collecting the relative calibration data along with absolute time information at a trigger rate of  $\sim 400$ Hz. Other details regarding hardware and software features of the data acquisition and control system of the telescope are discussed in [2.11].

A cost-effective method for operating the imaging camera of the TACTIC  $\gamma$ -ray telescope at stable single channel rates (SCR) and safe anode current values is being used despite variations in the light of the night sky experienced by the individual pixels from time to time [2.12]. The camera operates 13 PMTs ( 9 in the central region and 4 in the periphery of the camera) with fixed high voltages and the remaining 336 pixels at different high voltages to ensure their operation within a pre-determined Single Channel Rate (SCR) range. The purpose behind using the central 9 pixels of the camera at fixed high voltages is to facilitate the gain normalization (flat fielding) of the remaining 336 camera pixels, so that the event sizes ( $\equiv$  sum of CDC counts in the clean, flat-fielded image) recorded during a nights observations can be directly compared to one another. Operation of the pixels in a narrow SCR band has the advantage of ensuring a stable chance coincidence rate which can be used as a system diagnostic parameter. An elaborate algorithm [2.12] has been developed to monitor the SCR rates of all pixels using the CAMAC front ends and ensure their operation within a narrow range despite changes in the background light level incident on them due to changes in the sky brightness and

star-field rotation. The algorithm also ensures that all the pixels of the camera operate within safe anode current ranges. The feedback loop of the algorithm changes the high voltage to the various pixels using multichannel high voltage unit which has a resolution of 1V. The decision of operating a pixel under enhanced light levels is solely based on the comparison of the SCR and applied high voltage with reference data generated under controlled light level conditions. A detailed description of the single channel rate stabilization scheme can be found in [2.12].

## 2.6 Relative and absolute gain calibration scheme

The PMT calibration scheme employed for TACTIC has two parts, viz., relative gain calibration and absolute gain calibration. In the relative gain calibration scheme, we use a high intensity blue LED ( Nichia Japan make SPB 500) at a distance of  $\sim 2$ m from the camera to determine the relative gain of the imaging camera pixels. The LED has been provided with a light-diffusing medium in front of it to ensure the uniformity of its photon field within  $\sim \pm 6\%$ . The mean light intensity from the pulsed LED recorded by each pixel, in response to 2000 light flashes is subsequently used for off-line relative gain calibration of the imaging camera.

The absolute gain calibration system of the camera involves monitoring the absolute gain of a set of 4 gain calibrated pixels placed at the periphery of the camera. Since measurement of the absolute gains of these PMTs by determining their single photoelectron peaks a number of times during an observation run is rather time consuming, we have instead used a relatively simpler method of measuring the light pulser yield of a calibrated source for the in-situ determination of the absolute gain of these calibration channels. The calibrated light sources used are the  $\text{Am}^{241}$ -based light pulsers (Scionix Holland BV make; dimensions of YAP:Ce pulser units - 4 mm $\times$ 1 mm) which produce fast optical flashes at an average rate of  $\sim 20$  Hz with maximum emission at a wavelength of  $\sim 370$ nm. After taking several measurements of single photoelectron peak and the radioactive light pulser (RLP) yield under dark room conditions to validate the reproducibility of the measurements and for preparing the reference data base, the PMTs of these calibration channels are mounted permanently with radioactive light pulsers. A collimator is also used during the mounting of a light pulser on a particular tube so that the number of photoelectrons per pulse is  $\sim 500$  pe. These calibration pixels are operated at fixed high voltage values and any changes in the RLP yield measured during actual observations, are then attributed to actual gain change of the pixels. Since the 4 calibration pixels are also partially exposed to the light flashes from the LED during the relative calibration run, it becomes possible to determine the gain of all the pixels of the camera. A representative example of single photoelectron response of one of the PMT's is shown in Fig 2.5a. The mean amplitude of the single photoelectron peak (indicated by  $A_{SPE}$  in Fig 2.5a) is then determined by fitting a Gaussian distribution function to the differential rate curve. Fig 2.5b gives the pulse height

distribution of the light flashes obtained with one of the radioactive light pulsers. The underlying principle for converting the charge content of an uncalibrated pixel

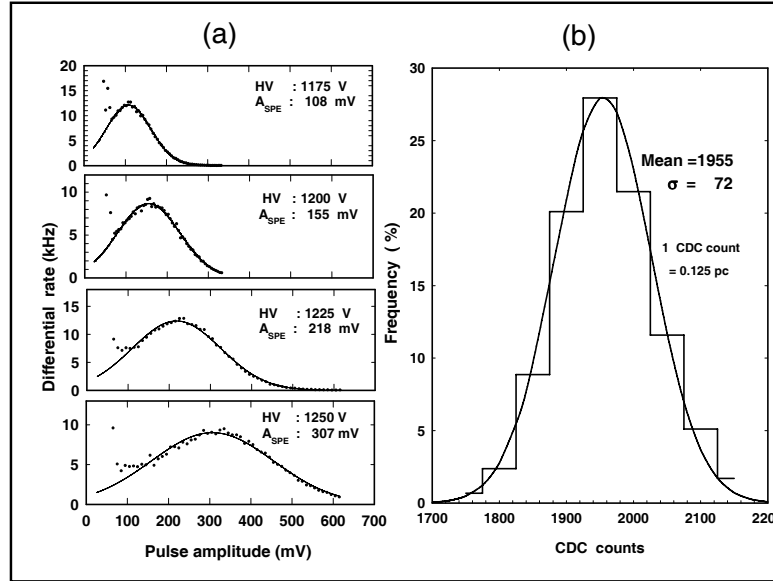


Figure 2.5: (a) An example of single photoelectron peak obtained for one of the calibration pixels at different values of HV. (b) Representative example of the pulse height distribution obtained with an Am<sup>241</sup>-based light pulser.

from CDC counts to photoelectrons, uses the fact that the calibration pixels are also exposed to the light flashes from the LED during the relative calibration run and hence it becomes possible to obtain the conversion factors for all the remaining 345 pixels of the camera [2.13]. The conversion for image size in CDC counts to number of photoelectrons has also been performed independently by using the excess noise factor method. The analysis of relative calibration data yields a value of  $1\text{pe} \cong (6.5 \pm 1.2)$  CDC for this conversion factor when an average value of  $\sim 1.7$  is used for excess noise factor of the photomultiplier tubes.

## 2.7 Monte Carlo simulations

Due to the non-availability of a calibrated beam of very high energy  $\gamma$ -ray photons, detailed Monte Carlo simulations offer the only way to benchmark the design and performance of an atmospheric Cherenkov imaging telescope. Measurements of absolute  $\gamma$ -ray flux and energy spectra of established  $\gamma$ -ray sources, as well as determination of upper limits on  $\gamma$ -ray emission from quiet objects also rely heavily on Monte Carlo predictions. We have used the CORSIKA (version 5.6211) air shower simulation code [2.14] for predicting and optimizing the performance of the TAC-TIC imaging telescope. The complete execution of the Monte Carlo simulations for

TACTIC telescope was subdivided into two steps. The first part comprised generating the air showers induced by different primaries and recording the relevant raw Cherenkov data (data base generation). Folding in the light collector characteristics and PMT detector response was performed in the second part. The simulated data-base for  $\gamma$ -ray showers used about 34000 showers in the energy range 0.2-20TeV with an impact parameter of 5-250m. These showers have been generated at 5 different zenith angles ( $\theta = 5^\circ, 15^\circ, 25^\circ, 35^\circ$  and  $45^\circ$ ). A data-base of about 39000 proton initiated showers in the energy range 0.4-40TeV, were used for studying the gamma/hadron separation capability of the telescope. The incidence angle of the proton showers was simulated by randomizing the shower directions in a field of view of  $6^\circ \times 6^\circ$  around the pointing direction of the telescope. The Cherenkov photons are ray-traced to the detector focal plane and the number of photoelectrons (pe) likely to be registered in a PMT pixel are inferred after folding in the relevant optical characteristics of the mirrors, the metallic compound-paraboloid light concentrator at the entrance window of the pixels and the photocathode spectral response. The Cherenkov photon data-base, consisting of number of photoelectrons registered by each pixel, is then subjected to noise injection, trigger condition check and image cleaning. The resulting data-bases, consisting of pe distribution in the imaging camera at various core distances and zenith angles are then used for estimating (a) trigger efficiency (b) effective detection area (c) optimum ranges of Cherenkov image parameters for discriminating between  $\gamma$ -ray and cosmic-ray events (d) differential count rate for  $\gamma$ -ray and cosmic-ray events and (e) effective threshold energy of the telescope for  $\gamma$ -ray and cosmic-ray proton events. The clean Cherenkov images were characterized by calculating their standard image parameters (defined in chapter 1) LENGTH, WIDTH, DISTANCE,  $\alpha$ , SIZE and FRAC2 [2.15,2.16]. The standard Dynamic Supercuts [2.17] procedure was used to separate  $\gamma$ -ray like images from the background cosmic-rays. In this method the image parameters listed above are adjusted to maximise the  $\gamma$ -ray signal in comparison to the background retention due to the cosmic-rays. The optimized values for these imaging parameters used are :

$$\begin{aligned}
0.11^\circ &\leq LENGTH \leq (0.260 + 0.0265 \times \ln S)^\circ, \\
0.06^\circ &\leq WIDTH \leq (0.110 + 0.0120 \times \ln S)^\circ, \\
0.52^\circ &\leq DISTANCE \leq 1.27^\circ \cos^{0.88} \theta, \\
SIZE &\geq 450 d.c \text{ ( where } 6.5 \text{ digital counts} \equiv 1.0 \text{ pe )}, \\
\alpha &\leq 18^\circ \text{ and} \\
FRAC2 &\geq 0.35.
\end{aligned}$$

The effective collection area of the telescope, for  $\gamma$ -ray and proton events, at two representative zenith angles values of  $15^\circ$  and  $35^\circ$  is shown in Fig 2.6a and Fig 2.6b, respectively.

These results were obtained by using the nearest neighbour topological-trigger with  $11 \times 11$  trigger field and a single pixel threshold of  $\geq 25$  pe. Fig 2.6c and



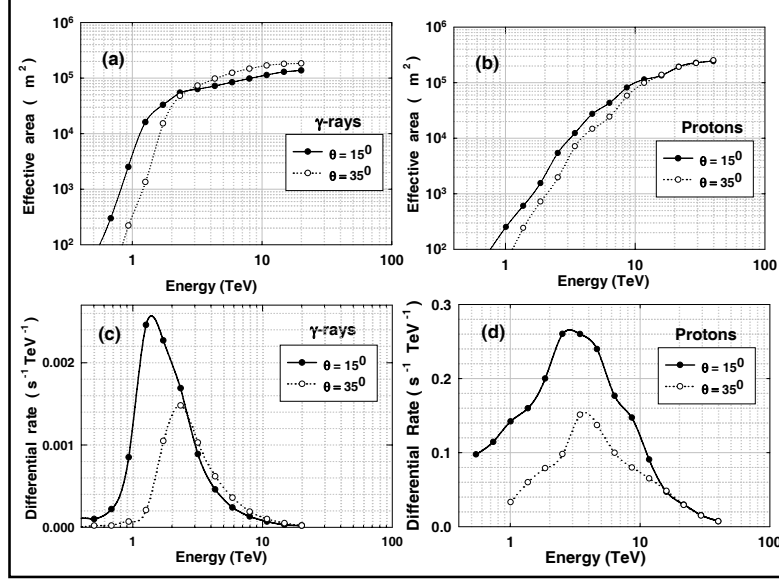


Figure 2.6: Effective collection area of the telescope for (a)  $\gamma$ -rays and (b) protons as a function of the primary energy at zenith angles of  $15^\circ$  and  $35^\circ$ . Differential trigger rates for (c)  $\gamma$ -rays and (d) proton events as a function of the primary energy.

2.6d show the corresponding differential event rates as a function of the primary energy for  $\gamma$ -ray and proton events, respectively. Defined as the energy where the differential rate peaks and assuming a Crab Nebula type of spectrum with a differential exponent of  $\sim -2.62$  [2.18], it is evident from Fig 2.6c that the  $\gamma$ -ray trigger threshold energy of the telescope is  $\sim 1.2\text{TeV}$ . The corresponding trigger threshold energy of the telescope for protons turns out to be to  $\sim 2.5\text{TeV}$  (Fig 2.6d.)

### 2.7.1 Comparison with real data

The agreement between the predictions from Monte Carlo simulations and the actual performance of the telescope was first checked by comparing the observed trigger rate of the telescope with the predicted value. The expected prompt coincidence rate at a zenith angle of  $15^\circ$  turns out to be  $\sim 2.5$  Hz for the nearest neighbour pair trigger mode. This value has been obtained on the basis of integrating the differential rate curve for protons (Fig 2.6d). Reasonably good matching of this with the experimentally observed value of  $\sim 2\text{-}3$  Hz suggests that the response of the telescope is very close to that predicted by simulations. A representative example of the behaviour of prompt coincidence rate as a function of zenith angle is shown in Fig 2.7. The result shown in Fig 2.7, is the analysis of the NN-2 trigger mode ON-0506 data collected for  $\sim 101.04$  h. The total data in this spell has been

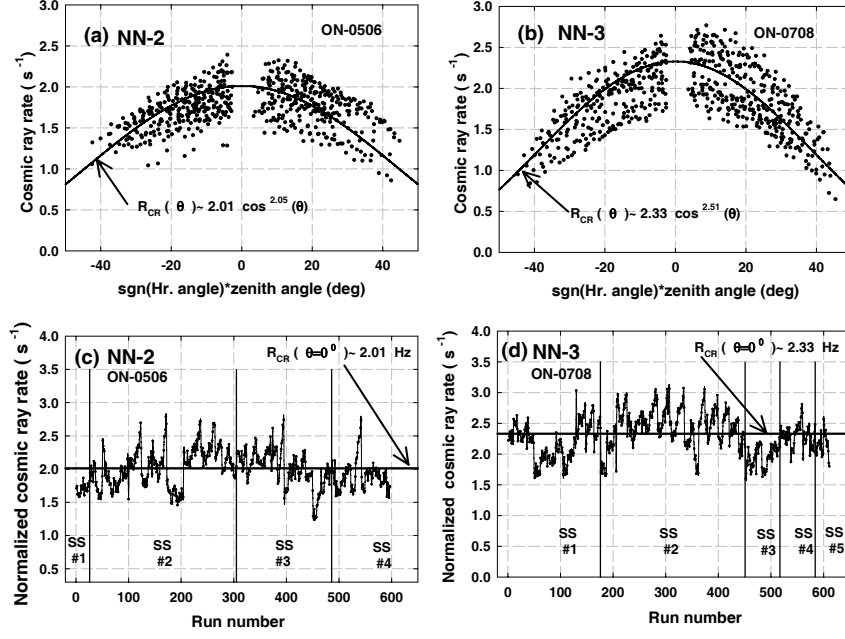


Figure 2.7: Measured cosmic ray trigger rate as a function of zenith angle (a) ON-0506 spell collected with NN-2 trigger configuration (b) ON-0708 spell collected with NN-3 trigger configuration. Normalized cosmic-ray rate as a function of run number (c) ON-0506 spell (d) ON-0708 spell. The segmentation of the data into various sub spells, on the basis of one lunation period, are indicated by full vertical lines.

divided into sub spells, where each sub spell corresponds to one lunation period. Each point in this plot represents a 10 minute observation run and the event rate has been calculated by dividing the number of events recorded in 600s. The behavior of the measured values of cosmic-ray rate as a function of zenith angle for the ON-0708 data taken with NN-3 trigger mode is also shown in Fig 2.7. In this case, the total data has been divided into 5 sub spells. The corresponding Monte Carlo estimates (solid line), with  $R_{CR}(\theta) \sim 2.01(\cos\theta)^{2.05} \text{ Hz}$  for NN-2 trigger and  $R_{CR}(\theta) \sim 2.33(\cos\theta)^{2.51} \text{ Hz}$  for NN-3 trigger, are also shown in figure. Detailed results of Monte Carlo simulations of  $\gamma$ -ray and cosmic ray proton induced extensive air showers as detected by the TACTIC atmospheric Cherenkov imaging telescope for optimizing its trigger field of view and topological trigger generation scheme can be found in [2.19].

The agreement between the expected and actual performance of the telescope was next checked by comparing the expected and observed image parameter distributions. Fig 2.8 shows the distributions of the image parameters LENGTH, WIDTH, DISTANCE and  $\alpha$  for simulated protons and for the actual Cherenkov

images recorded by the telescope. The simulated distributions of these image parameters for  $\gamma$ -rays have also been shown in Fig 2.8 for comparison. The observed

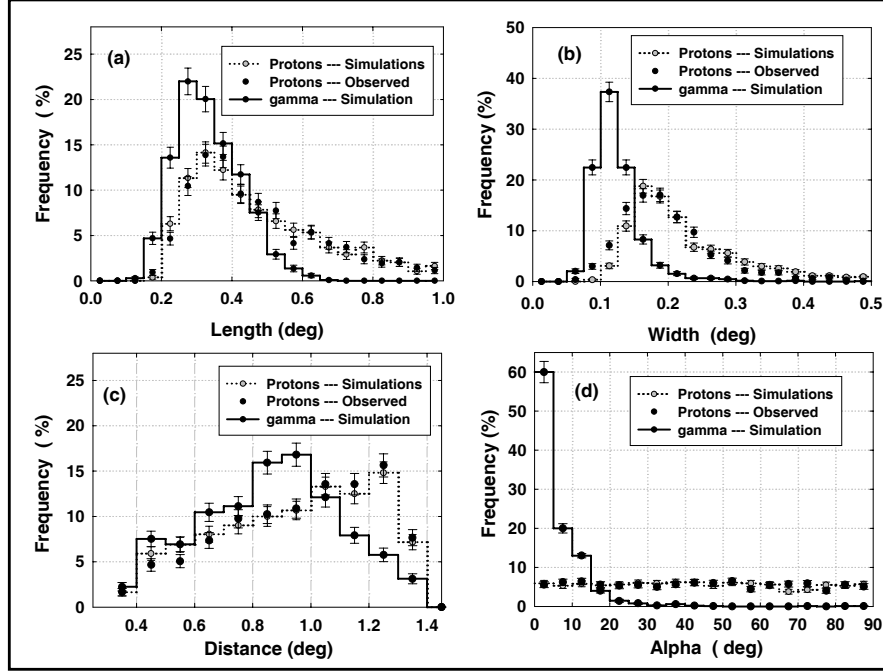


Figure 2.8: Comparison of image parameter distributions (a) LENGTH, (b) WIDTH, (c) DISTANCE and (d) ALPHA from real and the Monte Carlo simulated data for proton events. The simulated image parameter distribution of  $\gamma$ -rays has also been shown in the figure for comparison.

image parameter distributions are found to closely match the distributions obtained from simulations for proton-initiated showers, testifying to the fact that the event triggers are dominated by background cosmic-rays.

## 2.8 Performance evaluation using Crab Nebula observations

TACTIC telescope with a prototype camera of 81 pixels has been successful in detecting intense TeV  $\gamma$ -ray flaring activity from the BL-Lac object Mkn-501. This detection has an important historical significance in the field of very high energy  $\gamma$ -ray astronomy for being the first ever observation of a TeV  $\gamma$ -ray source by 5 independent groups [2.20]. Since 1997, the TACTIC imaging telescope camera and its data acquisition system has been continuously upgraded. Observations on potential  $\gamma$ -ray sources, however, were continued during this interim period of upgradation phase whenever it became possible. It was during Dec 2000-Mar 2001 that the

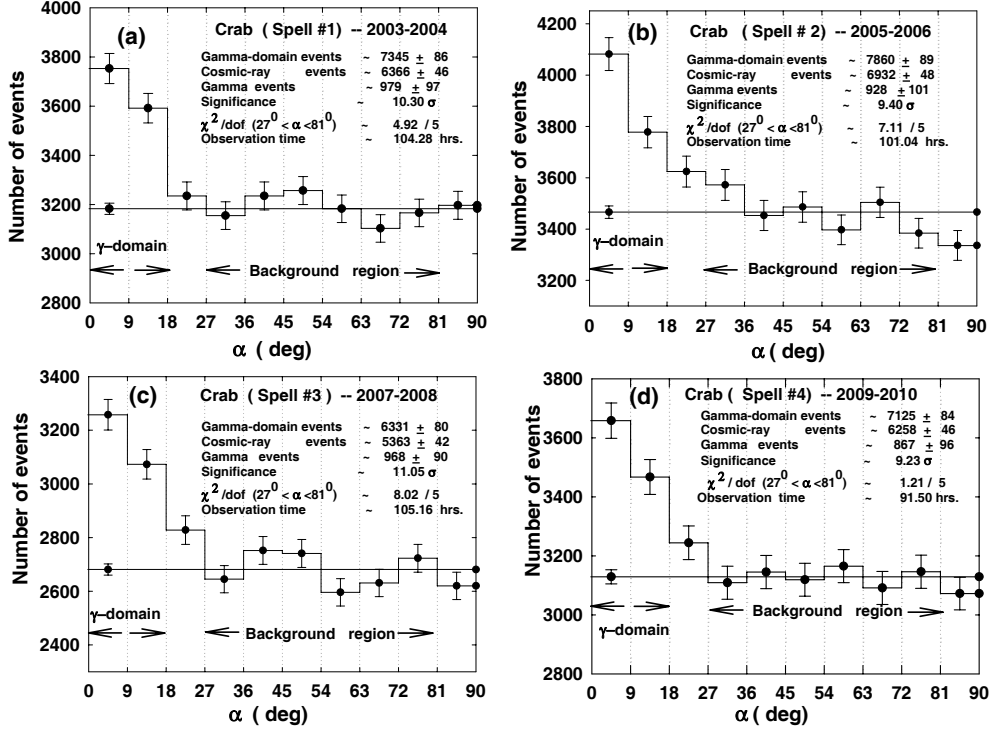


Figure 2.9: (a) Alpha plot for the Crab Nebula when ON-0304 on-source data is analyzed. (b), (c) and (d) same as (a) except for the ON-0506, ON-0708 and ON-0810 on-source data spells.

TACTIC imaging telescope in its full configuration of 349 pixels was able to detect  $\gamma$ -rays from the Crab Nebula and another BL-Lac object Mkn-421 at high statistical significances [2.21]. The results of the Crab Nebula observations, carried between the period 2003-2010 are shown in Fig 2.9 (a-d). Referring to this figure, one can clearly see an evidence of a  $\gamma$ -ray signal when Dynamic supercuts are applied to the data. The number of  $\gamma$ -ray events for the 4 spells shown in figure are determined to be  $\sim (979 \pm 97)$ ,  $\sim (928 \pm 101)$ ,  $\sim (968 \pm 90)$  and  $\sim (867 \pm 96)$  with corresponding statistical significances of  $\sim 10.3\sigma$ ,  $\sim 9.4\sigma$ ,  $\sim 11.05\sigma$  and  $\sim 9.23\sigma$ , respectively. The resulting  $\gamma$ -ray rates for the 4 on-source data spells turn out to be  $\sim (9.38 \pm 0.93)\text{h}^{-1}$ ,  $\sim (9.40 \pm 1.00)\text{h}^{-1}$ ,  $\sim (9.21 \pm 0.86)\text{h}^{-1}$  and  $\sim (9.47 \pm 1.05)\text{h}^{-1}$ , respectively. The values of  $\chi^2/\text{dof}$  for the background region (also shown in Fig 2.9) and the corresponding probability is also consistent with the assumption that the background region is flat and thus can be reliably used for estimating the background level in the  $\gamma$ -domain. Combining all the data for this period, the total data yields an excess of  $\sim (3742 \pm 192)$   $\gamma$ -ray events with a statistical significance of  $\sim 19.92\sigma$ . The resulting  $\gamma$ -ray rate turns out to be  $\sim (9.31 \pm 0.48)\text{h}^{-1}$ . The value of

$\chi^2/\text{dof} \sim 5.57/5$  for the the background region with the corresponding probability of  $\sim 0.35$  is again consistent with the assumption that the background region is flat. The aim of presenting these plots is to validate that the observed excess events are indeed genuine  $\gamma$ -rays from the source. On the basis of these results one can confidently say that the TACTIC telescope has consistently detected  $\gamma$ -ray emission from the ‘standard candle’ Crab Nebula during the 4 observations spells at a sensitivity level  $N_\sigma \sim 0.98 \sqrt{T}$ , where T is the observation time in hours. The corresponding number of  $\gamma$ -rays recorded is seen to follow  $N_\gamma \sim 9.16 T$ .

## 2.9 Sensitivity estimates

In this section we shall estimate the sensitivity of the TACTIC telescope. Assuming that main background against which  $\gamma$ -ray signal needs to be detected is due to protons and the detection sensitivity is limited by statistical fluctuations of  $\gamma$ -ray domain events (signal+background) and background region events, the expression for the statistical significance ( $N_\sigma$ ), using Equation 9 of Li and Ma [2.22], is given by :

$$N_\sigma = \frac{(N_{on} - \beta_1 N_{off})}{\sqrt{\beta_1(N_{on} + N_{off})}} = \frac{(N_\gamma)}{\sqrt{\beta_1(N_{on} + N_{off})}} \quad (2.1)$$

where  $N_{on}$  is the number of events in the  $\gamma$ -ray domain ( $\alpha \leq 18^\circ$ ),  $N_{off}$  is the number of events in the background region ( $27^\circ \leq \alpha \leq 81^\circ$ ) and  $\beta_1 (=2/6)$  is the ratio of  $\gamma$ -ray domain ( $\alpha \leq 18^\circ$ ) to background region ( $27^\circ \leq \alpha \leq 81^\circ$ ). Using the fact that only 60% (denoted by  $\beta_2$  where  $\beta_2=0.6$ ) of the total cosmic-ray events contribute to the  $\alpha$ -range of the background region with  $27^\circ \leq \alpha \leq 81^\circ$ , we get  $N_{on} = N_{\gamma 0} + \beta_1 \beta_2 N_{p0}$  and  $N_{off} = \beta_2 N_{p0}$  ( $N_{\gamma 0}$  and  $N_{p0}$  are the number of  $\gamma$ -ray and cosmic-ray events, respectively without applying any  $\gamma$ -domain cuts).

One of the ways of estimating the sensitivity is to calculate the minimum signal recovery time needed for detecting a  $\gamma$ -ray signal at an adequately high statistical significance (usually  $N_\sigma=5$   $\sigma$  is used as a benchmark in the field). Using the definition  $N_\sigma$  of given in equation 2.1, the expression for the minimum signal recovery time ( $T_{min}$ ) is given by :

$$T_{min} \approx N_\sigma^2 \left[ \frac{\beta_1}{R_\gamma} + (\beta_1(\beta_1 + 1)\beta_2) \frac{R_p}{R_\gamma^2} \right] \quad (2.2)$$

where  $R_\gamma$  and  $R_p$  are the  $\gamma$ -ray and cosmic-ray rates surviving the  $\gamma$ -domain cuts, respectively and these are given by:

$$R_\gamma = \int_{E_{min}}^{E_{max}} \frac{dF_\gamma}{dE} A_\gamma(E) \eta_\gamma(E) dE \quad (2.3)$$

$$R_p = \int_{E_{min}}^{E_{max}} \frac{dF_p}{dE} A_p(E) \eta_p(E) dE \quad (2.4)$$

where  $dF_\gamma/dE \sim 2.79 \times 10^{-7} (E/1\text{TeV})^{-2.57} m^{-2} s^{-1} \text{TeV}^{-1}$  is the differential energy spectrum of the Crab Nebula as measured by the HEGRA group [2.23],  $A_\gamma(E)$  is the effective collection area for  $\gamma$ -rays in  $m^2$ ,  $dF_p/dE \sim 0.096 E^{-2.7} m^{-2} s^{-1} sr^{-1} \text{TeV}^{-1}$  is the differential energy spectrum of the protons [2.24],  $A_p$  is the effective collection area for protons in units of  $m^2 sr$ ; and as already defined,  $\eta_\gamma$  and  $\eta_p$  are the  $\gamma$ -ray and protons acceptance factors, respectively after the application of selection cuts. Although many groups have used Quality Factor (QF) (where  $QF = \eta_\gamma / \sqrt{\eta_p}$ ) for estimating the sensitivity and for optimizing the performance of their classification methods, we will use  $T_{min}$  for estimating the performance of the telescope. The reason for this is the fact that a high value of QF can also result from tight cut which can reduce the  $\gamma$ -ray retention factor significantly.

It is evident from equation 2.2 that the calculation of  $T_{min}$ , for a particular  $\gamma$ -ray source, involves using effective collection areas for  $\gamma$ -rays and cosmic-rays and the corresponding retention factors after applying  $\gamma$ -domain cuts. Since all these variables, apart from being energy dependent are also a function of zenith angle, rigorous calculation of  $T_{min}$ , for a particular source, can be performed only if the actual zenith angle coverage is also duly considered in the calculation. While we have accounted for all the above mentioned dependences in determining the energy spectra of the Crab nebula, Mrk 421 and Mrk 501 [2.25-2.28], useful knowledge about the system performance can be also obtained if the calculations are performed on a basis of simple toy model. In this model, we use typical average values of these quantities so that one can directly find out the expected cosmic-ray and  $\gamma$ -ray rates by multiplying their integral spectra with effective collection areas. With these approximations expression 2.2 can be rewritten as :

$$T_{min} \approx N_\sigma^2 \left[ \frac{\beta_1}{R_{\gamma 0} \eta_\gamma} + (\beta_1(\beta_1 + 1)\beta_2) \frac{R_{p0} \eta_p}{R_{\gamma 0}^2 \eta_\gamma^2} \right] \quad (2.5)$$

where  $R_{\gamma 0}$  and  $R_{p0}$  are the  $\gamma$ -ray and cosmic ray rates without applying any  $\gamma$ -domain cuts, respectively. With  $QF = \eta_\gamma / \sqrt{\eta_p}$ , the above expression can be rewritten as

$$T_{min} \approx N_\sigma^2 \left[ \frac{\beta_1}{R_{\gamma 0} \eta_\gamma} + (\beta_1(\beta_1 + 1)\beta_2) \frac{R_{p0}}{R_{\gamma 0}^2 QF^2} \right] \quad (2.6)$$

Following this simplified approach, Fig 2.10 shows the sensitivity of the TACTIC telescope as a function of observed  $\gamma$ -ray rate (i.e  $R_\gamma$ ). The sensitivity plotted in Fig 2.10 is in terms of the signal recovery time for detecting a  $5\sigma$  steady signal from the Crab nebula at an energy above  $\sim 1.2\text{TeV}$  and by using  $A_\gamma \sim 4 \times 10^4 m^2$ . The estimated  $\gamma$ -ray rate without applying any  $\gamma$ -domain cuts is found to be  $\sim 19.1 h^{-1}$ . The corresponding  $\gamma$ -ray retention values have been also shown at the top of the figure so that one can estimate the  $\gamma$ -ray retention factor on the basis of the measured  $\gamma$ -ray rate. In order to estimate the cosmic-ray rate while one can follow a similar procedure, by using appropriate values of relevant parameters, the resulting predicted rate can be somewhat underestimated because of additional contribution

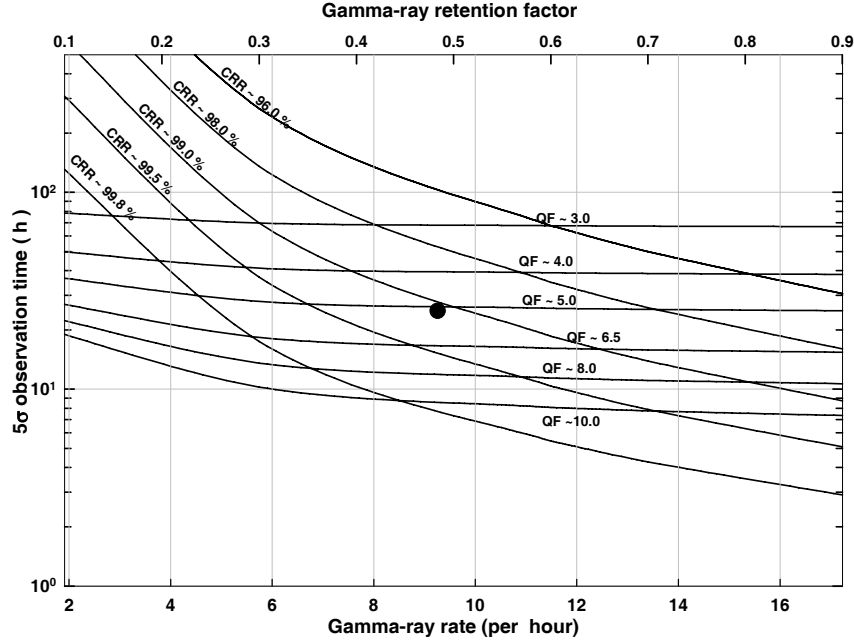


Figure 2.10: Sensitivity estimate of the TACTIC telescope in terms of the signal recovery time for  $5\sigma$  steady detection from Crab nebula as a function measured  $\gamma$ -ray rate above. Several representative curves obtained for different values of  $\eta_p$  and QF are also shown in the figure.

of  $\sim 30\%$  to the measured event rate due to other cosmic-ray nuclei [2.24]. While we have accounted for this appropriately in our detailed simulation work [2.19], for the simplified approach followed here, one can directly use a value of  $R_{p0} \sim 1.8\text{Hz}$  for estimating the cosmic ray rejection capability of the telescope. This value is obtained by taking the mean of observed cosmic-ray rates for the NN-2 and NN-3 trigger rates observed at a typical zenith angle of  $\sim 20^\circ$ .

On examination of Fig 2.10, it is evident that sensitivity level of  $\sim 5.0\sigma$  in  $\sim 25$  h (shown as filled circle in the figure) corresponds to  $\eta_\gamma \sim 0.50$  and  $\eta_p \sim 0.0093$  (i.e cosmic ray rejection  $\sim 99.07\%$ ). Several representative curves obtained for different values of  $\eta_p$  and QF are also shown in the figure so that useful guidance can be obtained from this figure for improving the sensitivity.

---

---

## 2.10 References

- [2.1] R. Koul et al., Nucl Instr and Meth A 578 (2007) 548.
- [2.2] R.K. Kaul et al., J.Astrophys. Astr., 15 (1994) 95.

- [2.3] S.K. Kaul et al., Meas. Sci. Technol., 8 (1997) 329.
- [2.4] V.K. Dhar et al., Meas. Sci. Technol., 21 (2010) 015112.
- [2.5] A.K. Tickoo et al., Exp. Astron., 9 (1999) 81.
- [2.6] A.K. Tickoo et al., Nucl. Instr. and Meth. A, 539 (2005) 177.
- [2.7] C.L. Bhat et al., Meas. Sci. Technol., 7 (1996) 706.
- [2.8] C.L. Bhat et al., Nucl. Instr. and Meth. A, 340 (1994) 413.
- [2.9] S.R. Kaul et al., Nucl. Instr. and Meth. A, 496 (2003) 400.
- [2.10] QNX System Architecture by QNX Software Systems Ltd., [www.qnx.com](http://www.qnx.com).
- [2.11] K.K. Yadav et al., Nucl. Instr. and Meth. A, 527( 2004) 411.
- [2.12] N. Bhatt et al., Meas. Sci. Technol., 12 (2001) 167.
- [2.13] A.K. Tickoo et al., Bull. Astr. Soc. India, 30 (2002) 381.
- [2.14] D. Heck et al., Report FZKA 6019 Forshungszentrum, Karlshruhe, (1998).
- [2.15] A.M. Hillas, Proc. 19th ICRC , La Jolla, 3 (1985) 445.
- [2.16] T.C. Weekes et al., Ap.J 342 (1989) 379.
- [2.17] G. Mohanty et al., Astropart. Phys., 9 (1998) 15.
- [2.18] F. Aharonian et al., Ap.J 614 (2004) 897.
- [2.19] M.K. Koul et al., Nucl. Instr. and Meth. A646 (2011) 204.
- [2.20] R.J.Protheroe et al., Proc. 25th ICRC, Kruger Park, South Africa 8 (1997) 317.
- [2.21] K.K. Yadav et al., Astropart. Phys. 27 (2007) 447.
- [2.22] T.P. Li and Y.Q.Ma, Ap.J 272 (1983) 317.
- [2.23] F. Aharonian et al., Ap.J 614 (2004) 897.
- [2.24] A. Konopelko, Astropart. Phys. 24 (2005) 191.
- [2.25] K.K. Yadav et al., Astropart. Phys. 27 (2007) 447.
- [2.26] P. Chandra et al., J.Phys. G : Nucl. Part. Phys. 37 (2010) 125201.
- [2.27] P. Chandra et al., J.Phys. G : Nucl. Part. Phys. 39 (2012) 045201.
- [2.28] S.V. Godambe et al., J.Phys. G : Nucl. Part. Phys. 35 (2008) 065202.



# Chapter 3

## Artificial Neural Network (ANN) Methodology

### 3.1 Introduction

The simulation of human intelligence using machines still remains a challenge to ingenuity. In the middle of the last century a research discipline namely the Artificial Intelligence (AI) emerged. The definition of the term AI is very distinct; a major reason for this is the fact that there is no commonly accepted definition for ‘intelligence’. The most comprehensive definition for AI includes all research methodologies aimed to simulate the intelligent behaviour. Thus it is safe to conclude that neural network theory grew out of Artificial Intelligence research, or from the research effort for designing machines with cognitive ability. It is therefore a computer program or a hardwired machine that is designed to learn in a manner similar to the human brain.

Artificial Neural Network (ANN) [3.1] is a field of major research interest at present, involving researchers from many disciplines. Subjects contributing to this research include biology, computing, electronics, mathematics, medicine, physics, astronomy, psychology etc. The approach to this interesting field of ANN is as diverse as its aims. The basic idea is to use the knowledge of nervous system and the human brain to design artificial intelligence systems.

While this quest is still in its infancy, biologists and psychologists are trying to model and understand the functionality of the nervous system and search for explanations for human behaviour. Simultaneously, scientists and engineers are searching for efficient ways to solve problems for which conventional methods are currently used.

### 3.1.1 Biological and Artificial Neural Networks (ANN)

ANN's draw much of their inspiration from biological nervous system. It is therefore essential to have some knowledge of the manner in which this system is organized.

Living creatures which have the ability to adapt to the changing environment, need a controlling unit which is able to learn. Higher developed animals and humans use very complex networks of highly specialized neurons to perform this task. The control unit, i.e, the brain can be divided into different sub-units, each having certain tasks like vision, hearing, motor and sensor control. The brain is connected by nerve sensors to the rest of body. The brain consists of a very large number of neurons, about  $10^{11}$  on an average. These can be seen as the basic building bricks for the central nervous system. The neurons are interconnected at points called synapses. The complexity of the brain is due to the massive number of highly interconnected simple units working in parallel, with an individual neuron receiving input from  $\sim 10^4$  others. The neuron contains all structures of an animal cell. The complexity of these structures and of the processes involved in a simple cell are enormous. Thus, even the most sophisticated neuron models which we replicate in artificial neural networks seem comparatively simple.

Structurally the neuron can be divided in three major parts : the cell body (soma/nucleus), the dendrite and the axon [Fig 3.1a]. Input connection are made from the axons of other cells to the dendrites or directly to the body of the cell. There is only one axon per neuron. It is a single and long fibre, which transports the output signal of the cell as electrical impulses along its length. The end of the axon may divide in many branches, which are then connected to other cells. The branches have the function to fan out signal to many other inputs. The neurons thus basically perform the following function: all the inputs to the cell, which may vary by the strength of connection or the frequency of the incoming signal, are summed up. The input sum is processed by a threshold function and produces an output signal. The processing time of about 1 ms per cycle and transmission speed of the neurons of about 0.6 to 120 m s<sup>-1</sup>, is very slow compared to a modern computer.

The brain however, works in both a parallel and serial manner. For example a human can recognise a picture of another person in  $\sim 100$ ms. Given the processing time of  $\sim 1$  ms for an individual neuron this implies that a certain number of neurons (but less than 100) are involved in serial mode; whereas the complexity of the task is the evidence for a parallel processing, because even a simple recognition task can not be performed by such a small number of neurons.

Biological neural systems usually have a very high fault tolerance. Experiments with people with brain injuries have shown that damage of neurons upto a certain level does not necessarily influence the performance of the system, though tasks such as writing or speaking may have to be learned again, i.e, the person has to re-train. Fig 3.1b shows an artificial implementation of the biological neuron. This is represented by the inputs which are analogous to the dendrites, the summing part

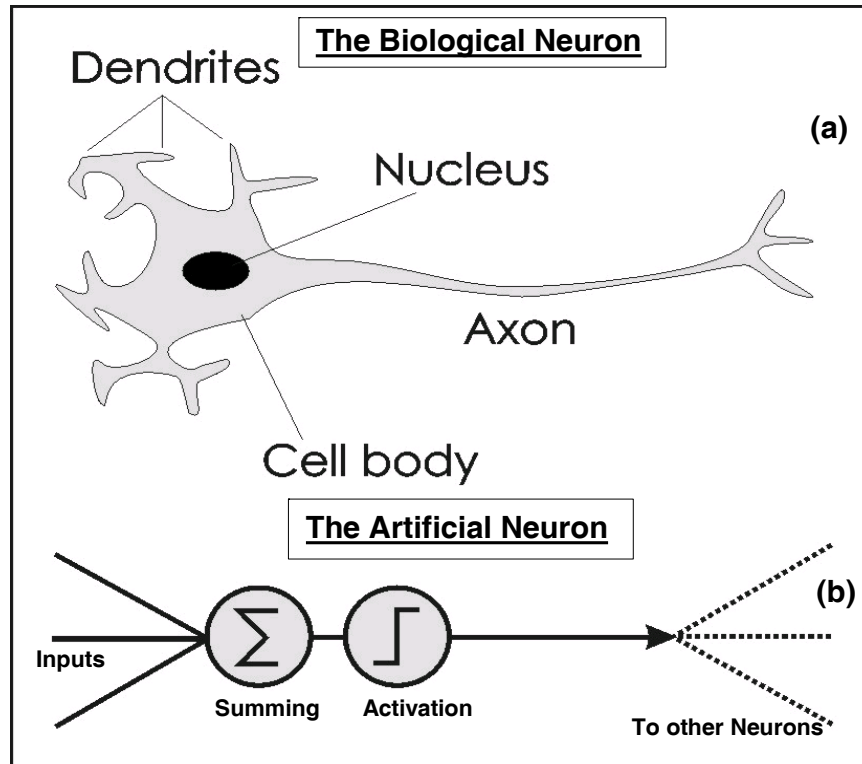


Figure 3.1: (a) Schematic diagram of a typical neuron or a nerve cell in the biological neuron. (b) The 'artificial' model of the biological neuron.

which is a rough equivalent of the synapse and the connection branching to other neurons which is an equivalent of the axon.

### Definitions of ANN

In literature a wide variety of definitions and explanations for the terms ANN can be found. The definition by Igor Aleksander [3.2] includes a wide range of methods and applications in the field of neural computing: 'Neural computing is the study of networks of adaptable nodes which, through a process of learning from task examples, store experimental knowledge and make it available for use'. According to Laurene Fausett [3.3], 'an ANN is an information-processing system that has certain performance characteristics in common with biological neural networks'. Robert Harvey [3.4] however focusses on the biological model. His definition excludes most parts of logical neural networks from the field of ANN. 'A neural network is a dynamical system with one interconnection. It carries out processing by its response to inputs'. There are many other definitions available in literature. In brief ANNs, are information processing devices whose design is inspired by studies

of the brain and the neuron system structure. Thus ANNs aim at simulating the activities of the human brain.

### **Memorization and Generalization**

To simulate the behaviour of brain, its abilities related to memorization and generalization are essential. These are two basic properties of ANN. To memorize basically means to commit to memory, i.e to learn so as to remember. To generalize means to form general principles or conclusions from detailed facts, experience etc. Memorising, given facts, is an obvious task in learning. This can be done by storing the input samples explicitly or by identifying the concept behind the input data, and memorising their general rules. This ability to identify the rules, to generalise, allows the system to make prediction on the unknown data.

Despite not being a strictly logical approach, the process of reasoning from specific samples can be observed in human learning process. Generalization importantly removes the need to store a large number of input samples. Features common to the whole class need not be repeated for each sample, instead the system only needs to remember which features are part of a sample.

### **Working of ANN**

Over the past few decades the biological discoveries have enhanced our knowledge about the structure and functioning of the human brain. Since a neuron is the basic computing unit of the brain it is necessary to study the details of the entire structure of neurons. Many researchers have given various architectures of the neuron based on its accumulative information processing properties. The first model of the neuron was proposed by McCulloch and Pitt in 1943 [3.5]. Mathematically this model can be described as :

$$Y_0 = f(y) = \begin{cases} 1....if & y \geq T; \\ 0...otherwise \end{cases} \quad (3.1)$$

This means that output of the neuron is 1 if the weighted sum of all the inputs exceeds the given threshold level  $T$  and is 0 otherwise.

### **Architecture of a single Neuron**

The architecture of a neuron can be demonstrated by Fig 3.1b. It shows the summation symbol (called aggregation Function) and the Activation Function. Though summation is the most commonly used function, however other functions like product of the terms and Radial Basis functions can also be used. Apart from this there are about two dozen activation functions which can be used, the most common being, Sigmoid, Hyperbolic Tangent, Sine Function, Gaussian, Cauchy, Decaying

exponential, Logarithmic etc. In the first half the weighted summation (i.e, summation of all inputs multiplied by corresponding weight) of all the inputs is performed and then the net value is transferred through activation function to the output of the neuron. The weights (or synaptic weights) denote the strength of linkage of the two neurons, i.e how strongly the two neurons are associated with each other.

### **Architecture of Neural Network**

Using a set of artificial neurons described above, a neural network can be formed. The first layer is called the input layer and the last layer is called the output layer. The layer in between the input layer and the output layer is called the hidden layer. There can be more than one hidden layer and several hidden neurons in them according to the complexity of the problem. The number of input neurons and the output neurons are fixed for a particular problem. The only changeable parameter in any model is the number of the nodes in the hidden layer and thus the weights which play an important role in learning.

## **3.2 Learning**

A crucial property of neural networks is that it can learn the desired response from a set of examples in the domain. This is in contrast with other approaches of computing, where an algorithm or rules are used to store the knowledge.

The advantage of learning from examples is that there is no need to explicitly form a rule system for the task at hand. To extract rules from the knowledge in the domain implies that there is some expert interpretation. This process is often difficult, especially if the experts have different opinion on the problem. From an abstract point of view training an ANN can thus be seen as an automatic process of extracting rules from a data set.

There are three basic paradigms of neural learning, supervised, unsupervised, and the reinforced learning. Supervised learning is a process where both information about environment (e.g, the sensory stimuli) and the desired reaction of the system (e.g, the motor response) is given. It is analogous to human learning with a teacher who knows all answers. In the ANN context supervised learning is a process of generalizing the vector pairs, i.e., the input vector together with the desired output vector. In this learning method, however, it is often tricky to determine when the learning process should be terminated.

A variant of supervised learning is called reinforcement learning. In this method the required output is not provided; the response by teacher is only whether the calculated result is 'right' or 'wrong', or (YES/NO), the actual answer is not communicated during learning. Thus the error signal generated during reinforced training is binary.

Unsupervised learning works only on the input vectors, and the desired output

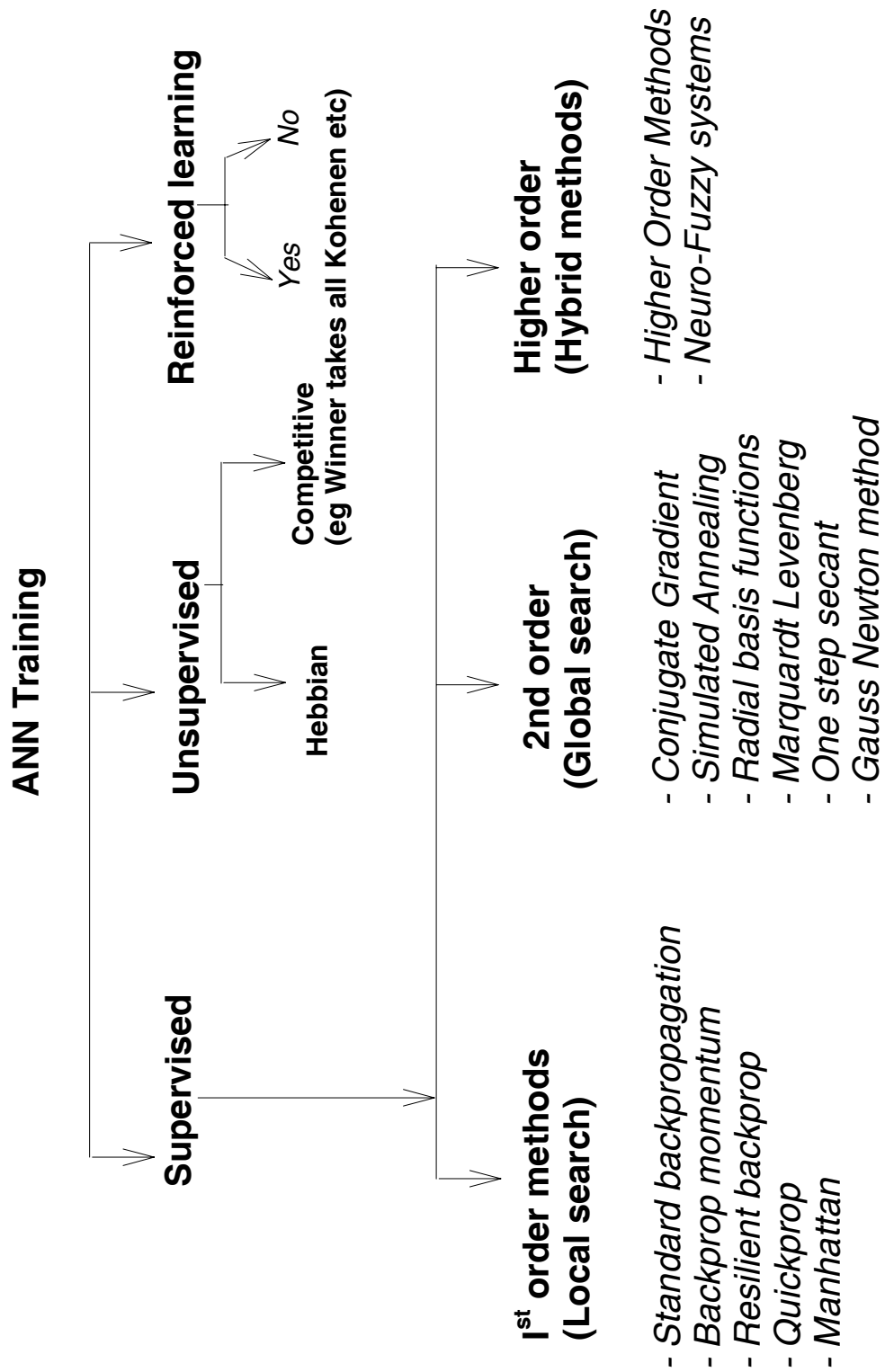


Figure 3.2: ANN classification based on the training schemes employed.

is not specified. This learning method can be compared to the process of categorization, discovering regularities, or adaptation to specific features. In many unsupervised models the categorization occurs according to the distance between the input vectors. Generalization based on this approach groups input vectors in a manner so as to minimize the distance between different members of a category for a given set. Unsupervised learning is closely related to the problem of density estimation in statistics. However unsupervised learning also encompasses many other techniques that seek to summarize and explain key features of the data. The generalized unsupervised learning can also broadly include methods like Clustering, blind signal separation, and dimensionality reduction methods like the Principal Component Analysis, Singular Value Decomposition etc. However in the context of Neural Networks, Self Organizing Maps (SOMs) (also referred to as the Kononen networks) and the Adaptive Resonance Theory (ART) are the commonly used unsupervised learning algorithms. We shall therefore restrict our discussion only to these methods.

The supervised and the unsupervised networks can be further classified as shown in Fig 3.2. The subdivision is based on the architecture in which the individual neurons are connected and the error minimization scheme adopted, thus there can be several possible ANN configurations.

### 3.3 Unsupervised methods

#### 3.3.1 Kohonen Self Organizing Maps (SOMs)

In such algorithms, no teacher is present and the algorithms have to self adapt to predict an answer. Learning and brain development of newborns is an important example which demonstrates this model. Let us consider how a newborn learns to focus its eyes, a skill not originally present, but acquired, inspite of the parent not being able to convey the technique to make sense of the visual stimuli impinging on the child's brain. However, after a few days, the child learns to associate sets with no help from outside. The possible answer to this was provided by Teuvo Kohonen [3.6]. The work provides a relatively fast and powerful model of how neural networks can self organize. In general, self organization refers to the ability of some networks to learn without being given a correct answer for the input pattern.

A Kohonen network is not a hierarchical system but consists of a fully interconnected array of neurons. The output of each neuron is an input to all other inputs in the network including itself. Each neuron has two sets of weights: one set is utilized to calculate sum of the weighted external inputs and the another one to control the interactions between different neurons in the network. The weights on the input pattern are adjustable, while the weights between neurons are fixed. The input is connected to all the nodes and there are interconnections between neurons of the same layer. During each presentation, the complete input pattern

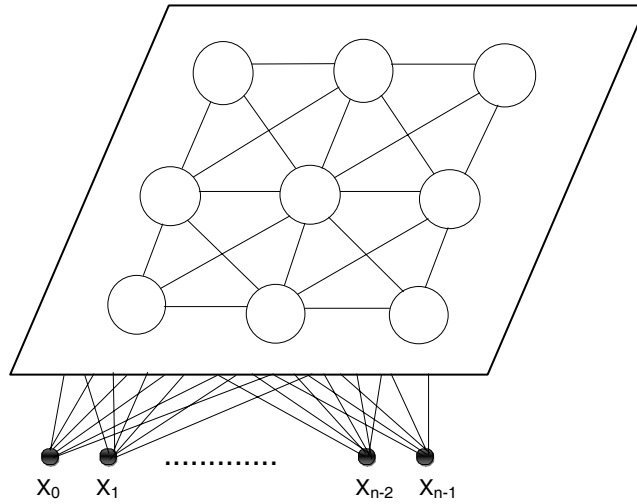


Figure 3.3: A two dimensional Kohonen network.

is presented to each neuron. The neuron computes its output as a sigmoidal function on the sum of its weighted inputs. The input pattern is then removed and the neurons interact with each other. The neuron with the largest activation function is declared as the winner and only this neuron is allowed to provide the output (*winner takes all method*) and its weight is updated and next iteration is computed. The block diagram of a Kohonen network is presented in Fig 3.3.

### 3.3.2 Adaptive Resonance Theory (ART)

The basic ART System is an unsupervised learning model. ART was developed by Carpenter and Grossberg[3.7]. The term ‘resonance’ used in ART, refers to resonant state of a neural network in which a category prototype vector matches close enough to the current input vector. ART matching leads to a resonant state, which permits learning. The network thus can be regarded as learning only in its resonant state. The key to ART is expectation, as inputs are presented, it is compared with a prototype vector that closely matches the expectation. If the matching is not adequate, a new prototype is presented. The ART network can adaptively create a new neuron corresponding to an input pattern if it is determined to be sufficiently different from existing clusters. This determination, called the vigilance test, is incorporated into the adaptive backward network. Thus the ART architecture allows the user to control the degree of similarity of patterns placed in the same cluster. Fig 3.4 shows the simplified ART architecture. The basic ART involves 3 groups of neurons, an input processing field ( $F_1$  layer) and the



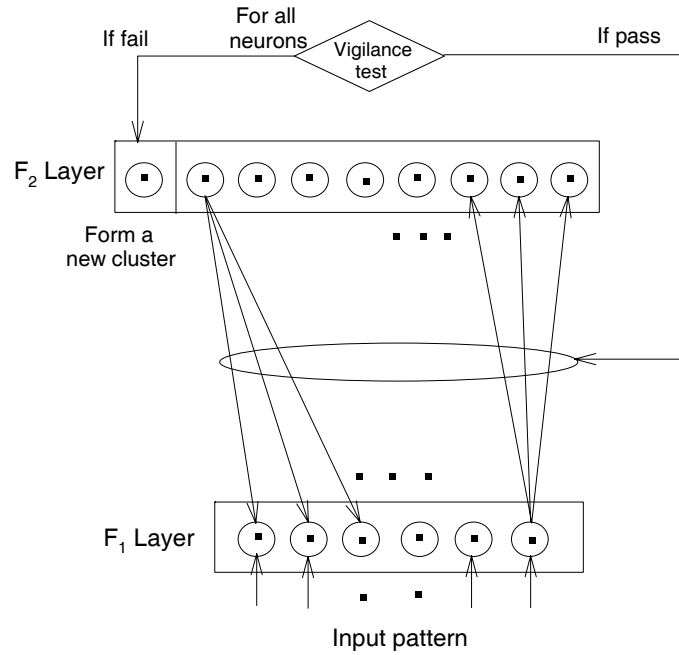


Figure 3.4: A simplified model of ART network.

cluster units ( $F_2$  layer) and a mechanism to control the degree of similarity of patterns placed on the same cluster (a reset mechanism). To control the similarity of patterns, there are two sets of connections (each with its own weights) between each unit of input of  $F_1$  layer and the cluster unit of the  $F_2$  layer. This setup is referred to as the bottom-up weights. Similarly  $F_2$  layer is connected to  $F_1$  layer by top down weights.

The  $F_2$  layer is a competitive layer. The cluster unit with the large net input becomes the candidate to learn the input pattern, setting all other  $F_2$  units to zero. The reset unit makes the decision whether the cluster unit is allowed to learn the input pattern depending on how similar its top-down weight vector is to the input vector. If the cluster unit is not allowed to learn (inhibited) then a new cluster unit is selected as the candidate. ART networks are further subdivided into ART1 and ART2. While as ART1 is a binary version of ART and can cluster binary input vectors, the ART 2 is an analogous version of ART. It can cluster real value input vectors.

### 3.3.3 Hebbian learning

This rule was proposed by D.Hebb in 1949 and is based on correlative weight adjustment. This is one of the oldest learning mechanisms inspired by biology. In

this the input-output pattern pairs  $(X_i, Y_i)$  are associated with the weight matrix  $\mathbf{W}$  known as the correlation matrix. It is computed as :

$$\mathbf{W} = \sum_{i=1}^n X_i Y_i^T$$

where  $Y_i^T$  is the transpose of the associated output vector  $Y_i$

### 3.3.4 Drawbacks of unsupervised networks

Despite being successful, unsupervised networks have not become very popular among the researchers. In astronomy, clustering algorithms have been used in past, to automatically divide data into many sub groups e.g., autoclass clustering program was used extensively to analyze infrared spectral data, which discovered a sub class of stars previously unknown to astronomers [3.8]. However these unsupervised algorithms are limited in what they can achieve. They can detect general trends in data but can not make use of additional knowledge to the problem at hand, which a human expert can. Also these methods use general notions to identify patterns in data but cannot specialize for specific problems. Further, as they are unguided methods, they may report patterns or trends which may not be of any interest, e.g., trends caused by some systematic error in data acquisition/analysis.

The alternatives to these are supervised learning methods which adapt to the problem at hand by not only exploiting but acquiring the problem specific knowledge. In the supervised methods, while algorithms like Standard Backpropagation (along with its variants like the Backprop-momentum, Quickprop etc) and the Resilient Backpropagation come under the category of 1st order algorithms, the Conjugate gradient methods, Levenberg-Marquardt algorithm, One Step Secant, Simulated Annealing, use of Radial Basis Functions, etc., belong to the category of 2nd order methods. Hybrid methods constitute the models like the Higher Order Neuron and the Neuro Fuzzy Systems. We shall in the next section present a brief description into these methods whose application has gained tremendous momentum in the recent years.

## 3.4 Supervised methods

### 3.4.1 First order algorithms

The conventional ANN which is based on McCulloch Pitts neuron model (Fig 3.1B) has been extensively used in many applications. A number of learning algorithms have been formulated to reduce the learning complexity of the ANN and also to achieve faster convergence. The McCulloch neuron model is a first order neuron model. It has serious limitations of mapping and classification and can not even represent a simple XOR problem. The first order learning schemes may be attributed

to the gradient decent method which is only an approximation of the truncated Taylor series.

### 3.4.2 Standard Backpropagation algorithm

In the standard backpropagation model, proposed by Rumelhart [3.9], a set of inputs is applied from outside, which are multiplied by weights, and the products are summed. This summation of products, denoted as (NET) is calculated for each neuron in the network. After NET is calculated, an activation function is applied to modify it, thereby producing the signal *OUT*. Sigmoid activation function of the type below is usually employed.

$$OUT = \frac{1}{1 + e^{-NET}} \quad (3.2)$$

The reason for employing the sigmoid activation function over many other non linear functions, is that it has a simpler derivative of the type :

$$\frac{\partial OUT}{\partial NET} = OUT(1 - OUT) \quad (3.3)$$

Also called as the ‘logistic’ or simply the ‘squashing function’ the sigmoid compresses the range of NET so that OUT lies between 0 and 1. Multilayer networks have greater representational power than single layer networks only if nonlinearity is introduced. The squashing function produces the desired nonlinearity.

There are many activation functions that can be used; the Backpropagation algorithm only requires that the function be nonlinear, differentiable and continuous everywhere. The advantage of using the sigmoid function, apart from a very simple derivative, is that it also provides a form of automatic gain control. For small signals (NET close to zero) the slope of the input/output curve is steep, producing a high gain. As the magnitude of the signal increases, the gain decreases. In this manner large signals can be accomodated by the network without saturation, while small signals are allowed to pass through without excessive attenuation. The literature is inconsistent in defining the number of layers needed in these networks. Some authors refer to the input as a layer, while others refer to the weights and the outputs only as a layer. Because the former definition is more functionally descriptive, it will be considered throughout this thesis. By definition, therefore, the above network is considered to consist of 3 layers. Backpropagation can be applied to networks with any number of layers; however only two layers of weights are needed to demonstrate the algorithm.

### 3.4.3 Overview of training

The objective of the training, is to adjust the weights so that a set of inputs produces the desired set of outputs. These input-output sets are also referred to as the

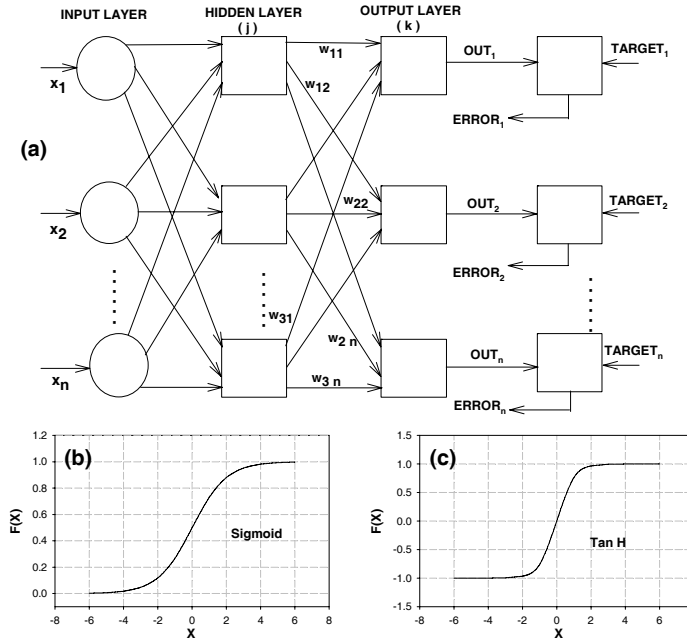


Figure 3.5: (a) A Feed Forward Backpropagation model with an input layer, 1 hidden layer and an output layer. (b) Sigmoidal Logistic function (c) Hyperbolic Tangent function, the two most commonly used activation functions.

‘vectors’. Training assumes that each input vector is paired with a target vector representing the desired output; together these are called a training pair. Usually, a network is trained over a number of training pairs. Before starting the training process all the weights must be initialized to small random numbers. This ensures that the network is not saturated by large value of weights and prevents certain other training pathologies, e.g if all the weights start at equal values and the desired performance requires unequal values, the network may not train.

Training of backpropagation requires the following steps:

1. Select the next training pair from the training set; apply the input vector to the network input.
2. Calculate the output of network.
3. Calculate the error between the network output and the desired output (target vector from training pair)
4. Adjust the weights of the network in a way that minimises the error.
5. Repeat the steps above for each vector in the training set until the error of the entire set is acceptably low.

The operations required in steps 1 and 2 above are similar to the way in which the trained network will ultimately be used. Calculations are performed on a layer by layer basis. Referring to Fig 3.5a above, first the outputs of neurons in the hidden layer (j) are calculated; these are then used as inputs to output layer (k), from which the network output values are calculated.

In step 3, each of the network outputs, labelled OUT in Fig 3.5b, is subtracted from its corresponding target vector to produce an error. This error is used in step 4 to adjust the weights of network, where the polarity and magnitude of the weight changes are determined by training algorithm. After enough repetitions of these four steps, the error between actual outputs and target outputs should be reduced to an acceptable value and the network is said to have been trained. At this point the network is used for recognition and the weights are not changed. It may be seen that first two steps constitute a ‘forward pass’ in that the signal propagates from the network input to the output. Last two steps maybe termed as the ‘reverse pass’ and the calculated error signal propagates backward through the network where it is used to adjust weights .

### Forward pass

First two steps can be expressed in vector form as follows: an input vector **X** is applied and an output vector **Y** is produced. The input target vector pair **X** and **T** come from the training set. The calculation is performed on **X** to produce the output vector **Y**. Calculation of multilayer networks is done layer by layer starting at the layer nearest to the inputs. The **NET** value of each neuron in the first layer is calculated as the weighted sum of its neuron inputs. The activation function *F* then “squashes” **NET** to produce the **OUT** value for each neuron in that layer. Once the set of outputs for a layer is found, it serves as input to the next layer. The process is repeated, layer by layer, until the final set of network outputs is produced. The weights between neurons can be considered to be a matrix **W**. Rather than using summation of products, the **NET** vector for the layer **N** maybe expressed as product of **X** and **W**. In vector notation **N**= **XW**. Applying the function *F* to the NET vector **N**, component by component, produces the output vector **O**. Thus for a given layer, the following expression describes the calculation process:

$$\mathbf{O} = F(\mathbf{XW}) \quad (3.4)$$

The output vector of one layer is the input vector for the next so calculating the outputs of the final layer requires the application of above equation to each layer, from the network’s input to its output.

### Reverse pass

In the reverse pass, three main steps are involved. (1) Adjusting the weights of

the output Layer. (2) Adjusting the weights of the hidden Layers and (3) Adding a neuron bias

Because a target value is available for each neuron in the output layer, adjusting the associated weights is easily accomplished using a modification of the delta rule. Delta rule is a method in which, given an input vector, the output vector is compared to the correct answer, if the difference is zero, no learning takes place, else, the weights are adjusted to reduce this difference. Interior layers are referred to as 'hidden layers', as their outputs have no target values for comparison, hence the training is far more complicated.

Assuming the training process for a single weight from neuron 'p' in hidden layer  $j$  to neuron 'q' in the output layer  $k$ . The output of neuron in layer  $k$  is subtracted from its target value to produce the ERROR signal. This is multiplied by the derivative of squashing function  $[OUT(1-OUT)]$  (from equation 3.3) calculated for the neuron layer  $k$ , thereby producing the  $\delta$  value.

$$\delta = OUT(1 - OUT)(Target - OUT) \quad (3.5)$$

Then  $\delta$  is multiplied by OUT from a neuron  $j$ , the source neuron for the weight in question. The product is in turn multiplied by a training rate coefficient  $\eta$  (typically 0.01 to 1.0) and the result is added to the weight. An identical process is performed for each weight proceeding from a neuron in the hidden layer to a neuron in the output layer employing the following formula:

$$\begin{aligned} \Delta W_{pq,k} &= \eta \delta_{q,k} OUT_{p,j} \\ W_{pq,k}(n+1) &= W_{pq,k}(n) + \Delta W_{pq,k} \end{aligned} \quad (3.6)$$

where  $W_{pq,k}(n)$  is the value of weight from neuron  $p$  in the hidden layer to the neuron  $q$  in the output layer at step  $n$  (before adjustment); subscript  $k$  indicates that the weight is associated with its destination layer,  $W_{pq,k}(n+1)$  is the value of the weight (after adjustment),  $\delta_{q,k}$  denotes the value of  $\delta$  for neuron  $q$  in the output layer  $k$  and  $OUT_{p,j}$  is the value of OUT for neuron  $p$  in the hidden layer  $j$ , subscripts  $p$  and  $q$  refer to specific neurons while as subscripts  $j$  and  $k$  refer to a layer. The procedure is illustrated in Fig 3.6a

Hidden layers have no target vector so that the above training process cannot be used. The hidden layers are trained by propagating the output error back through the network layer by layer, adjusting weights at each layer. The set of equations given at (3.6) are used for both the output and hidden nodes, however for hidden layer  $\delta$  must be generated without the benefit of a target vector.  $\delta$  is calculated for each neuron in the output layer, as in equation (3.5). It is used to adjust the weights feeding into the output layer, then it is propagated back through the same weights

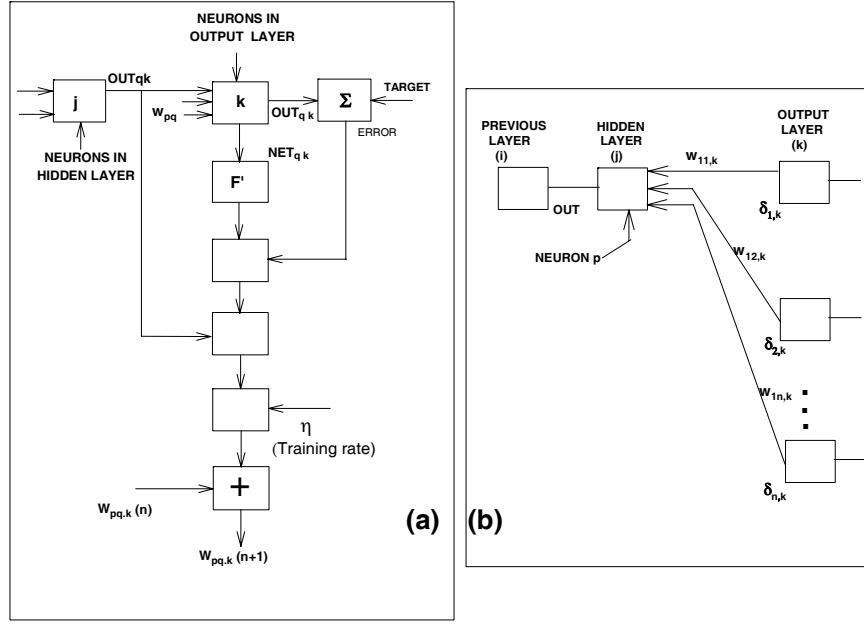


Figure 3.6: (a) Training of weights in the Output layer. (b) Training of weights in the Hidden layer where no target is available.

to generate a value of  $\delta$  for each neuron in the first hidden layer. These values of  $\delta$  are used in turn to adjust the weights of this hidden layer and in a similar way, are propagated back to all the preceding layers (Fig 3.6b).

Consider a single neuron in the hidden layer just before the output layer. In the forward pass, the neuron propagates its output value to neurons in the output layer through the interconnecting weights. During training these weights operate in reverse, passing a value  $\delta$  from the output layer back to the hidden layer. Each of these weights is multiplied by the  $\delta$  value of the neuron to which it connects in the output layer. The value of  $\delta$  needed for hidden layer neuron is produced by summing all such products and multiplying by the derivative of the squashing function.

$$\delta_{pj} = OUT_{pj}(1 - OUT_{pj}) \left( \sum_q \delta_{q,k} W_{pq,k} \right) \quad (3.7)$$

Knowing  $\delta$ , the weights feeding the first hidden layer can be adjusted using equation set (3.6) modifying indices to indicate the correct layers. For each neuron in a given hidden layer  $\delta$  must be calculated, and all weights associated with that layer must be adjusted. This is repeated, moving back towards the input layer by layer, until all weights are adjusted.

It is also desirable to provide each neuron with a trainable bias. This offsets the origin of logistic function, thereby permitting more rapid convergence of the training process. This feature is easily incorporated into training algorithm; a weight connected to +1 is added to each neuron. This weight is trainable in the same way as all the other weights, except that the source is always +1, instead of being the output of a neuron in a previous layer.

### Backprop-Momentum

This refers to addition of a term for improving the training of the backpropagation algorithm, while enhancing the stability of the process. Called as momentum, this method involves adding a term to the weight adjustment that is proportional to the amount of previous weight change. Once an adjustment is made it is ‘remembered’ and serves to modify all subsequent weight adjustments. The weight adjustment equations are modified to the following:

$$\begin{aligned}\Delta W_{pq,k}(n+1) &= \eta(\delta_{q,k}OUT_{p,j}) + \alpha[\Delta W_{pq,k}(n)] \\ W_{pq,k}(n+1) &= W_{pq,k}(n) + \Delta W_{pq,k}(n+1)\end{aligned}\tag{3.8}$$

where  $\alpha$  is the momentum coefficient and is commonly set around 0.9. Using the momentum method, the network tends to follow the bottom of narrow gullies in the error surface rather than crossing rapidly from side to side. This method, though works well on some problems, can however have little or sometimes even a negative effect on other applications.

#### 3.4.4 Caveats of the standard backpropagation model

Despite the most successful applications of backpropagation, the error minimization scheme followed by backpropagation method, called the gradient descent method, often converges very slowly. The success of this algorithm in solving large scale problems critically depends on user specified learning rate and momentum parameters and there are no standard guidelines for choosing these parameters. Some of the common caveats of the standard backpropagation algorithm are :

##### Network paralysis

As the network trains, the weights can get adjusted to very large values. This can force all or most of the neurons to operate at large values of OUT, in a region where derivative of activation function is very small. Since the error information to be sent back for training is proportional to its derivative, the training process can come to a virtual standstill. There is little theoretical undersanding of this problem. This can be avoided by reducing the step size  $\eta$ , which however extends the training time. Various methods have been employed to prevent paralysis, or



to recover from its effects, however a solid foundation for this is yet to be established.

### **Local minima**

Backpropagation employing gradient descent follows the slope of the error surface downwards constantly, adjusting the weights towards a minimum. The error surface of a complex network is highly convoluted, full of hills, valleys, folds and gullies in a multi-dimensional space. The network can get trapped in a local minima (a shallow valley) when there is a much deeper minimum nearby. From the limited viewpoint of the network, all directions are upward and hence it has no way to escape.

### **Step size**

In the backpropagation learning algorithm, small weight adjustments are assumed, which is clearly impractical, as it implies infinite training time. It is therefore necessary to select a finite step size. If the step size is too small, convergence can be very slow, however if the step size is large, paralysis or continuous instability can result.

### **Temporal instability**

If a network is learning to recognize e.g. alphabets, it does no good to learn B if, in doing so, it forgets A. A process is needed for teaching the network to learn an entire training set without disrupting what is already learned. The network should therefore be presented with all vectors in the training set before adjusting any weights. The needed weight change must be accumulated over the entire set, thereby requiring additional storage. After the number of such training cycles, the weight will converge to a minimal error. This method may not be useful if the network faces a continuously changing environment where it may never see the same input vector again. In this case, training process may never converge and wander aimlessly or oscillate widely. In this sense backpropagation fails to mimic biological systems.

## **3.4.5 Resilient backpropagation**

A major drawback of the gradient descent is the 'contra intuitive' influence of the partial derivative on the size of the weight-step. If the error function is shallow, the derivative is quite small, resulting in a small weight step. On the other hand, in the presence of steep ravines in the energy landscape, where cautious steps should be taken, large derivatives lead to large weight steps, possibly taking the algorithm to a completely different region of the weight space.

The basic principle of the resilient backpropagation algorithm (RPROP) [3.10],

is to eliminate the harmful influence of the size of partial derivative on the weight step. As a consequence, only the sign of derivative is considered to indicate the direction of weight update. The size of the weight change is exclusively determined by a weight specific, so called "update-value"  $\Delta_{ij}$ .

$$\Delta W_{ij} = \begin{cases} -\Delta_{ij}(t), \dots \text{if } \frac{\partial E}{\partial W_{ij}}(t) > 0 \\ +\Delta_{ij}(t), \dots \text{if } \frac{\partial E}{\partial W_{ij}}(t) < 0 \\ 0, \dots \text{otherwise.} \end{cases} \quad (3.9)$$

It is important to note here that replacing  $\Delta_{ij}$  by a constant update-value  $\Delta$  in (3.9) yields the so called **Manhattan learning**.

The second step of the RPROP learning is to determine the new update values  $\Delta_{ij}(t)$ . This is based on a sign dependant adaptation process, i.e

$$\Delta_{ij}(t) = \begin{cases} \eta^+ \times \Delta_{ij}(t-1), \dots \text{if } \frac{\partial E(t-1)}{\partial W_{ij}} \times \frac{\partial E(t)}{\partial W_{ij}} > 0 \\ \eta^- \times \Delta_{ij}(t-1), \dots \text{if } \frac{\partial E(t-1)}{\partial W_{ij}} \times \frac{\partial E(t)}{\partial W_{ij}} < 0 \\ 0, \dots \text{otherwise.} \end{cases} \quad (3.10)$$

where  $0 < \eta^- < 1 < \eta^+$ . At the begining all update values are set to the initial value  $\Delta_0$ , which is one of the two parameters of Rprop. Since  $\Delta_0$  directly determines the size of the first weight step, it should be chosen according to the initial value of the weights themselves, e.g,  $\Delta_0=0.1$ . The choice of this value is rather uncritical, for this is adapted as the learning progresses.

In order to prevent the weights from becoming too large, the maximum weight-step determined by the size of the update value is limited. The upper bound is set by the second parameter of Rprop,  $\Delta_{max}$ . The default upper bound is set arbitrarily to  $\Delta_{max}=50.0$ . Usually, convergence is rather insensitive to this parameter as well. Nevertheless, for some problems it can be advantageous to allow only very cautious/small steps, in order to prevent the algorithm getting stuck too quickly in suboptimal local minima.

The increase and the decrease factor are fixed to  $\eta^+=1.2$  and  $\eta^-=0.5$ . These values are based on both theoretical considerations and empirical evaluations. This reduces the number of free parameters to two, namely  $\Delta_0$  and  $\Delta_{max}$ .

Thus in contrast to the learning-rate based algorithms, Rprop modifies size of the weight-step directly by introducing the concept of resilient update-values. As a result the adaptation effort is not blurred by unforeseeable gradient behaviour. Due to clarity and simplicity of the learning laws, there is only a slight expense in computation compared with ordinary backpropagation.

However Rprop, too, suffers from certain problems common to the adaptive learning algorithms. As the adaptation is based on the estimation of the topology of the error function, both adaptation and weight update can be first performed after whole gradient information is available, i.e, after each pattern has been presented and the gradient of the sum of pattern errors is known. Accordingly, adaptive learning procedures are typically based on learning by epoch. This possibly reduces their efficiency on redundant training sets, compared to simple stochastic gradient descent and poses problems on their use with variable training sets.

Moreover, a restricted local adaptation inherently lacks the overall view that some 1<sup>st</sup> order (global) techniques may have. Hence, there is a need to study and adopt the more popular global search methods.

### 3.4.6 Quick-propagation algorithm

Standard back-propagation calculates the weight change based upon the first derivative of the error with respect to the weight. Quick-propagation method is a variation of standard back-propagation to speed up training [3.11]. This algorithm requires saving the previous gradient vector as well as previous weight change. The calculation of weight change uses only the information associated with the weight being updated.

$$\Delta W_{ij}(n) = \frac{\nabla W_{ij}(n)}{\nabla W_{ij}(n-1) - \nabla W_{ij}(n)} \times \Delta W_{ij}(n-1) \quad (3.11)$$

where  $\nabla W_{ij}(n)$  is the gradient vector component associated with weight vector  $W_{ij}$  in step  $n$ ,  $\nabla W_{ij}(n-1)$  is the gradient vector component associated with weight  $W_{ij}$  in the previous step and  $\Delta W_{ij}(n-1)$  is the weight change in step  $(n-1)$ . A maximum growth factor  $\mu$  is used to limit the rate of increase of step-size like :

If  $\Delta W_{ij}(n) > \mu \Delta W_{ij}(n-1)$ , then  $\Delta W_{ij}(n) = \mu \Delta W_{ij}(n-1)$ .  
Fahlman suggested an empirical value 1.75 for  $\mu$ .

There are however certain complications in this method also. First, the step-size calculation requires the previous value, which is not always available and has to be sometimes derived or assumed. Second, the weight values are unbounded, and can sometimes become so large that they overflow.

## 3.5 Second order algorithms

### 3.5.1 Conjugate gradient methods and its variants

The conjugate gradient method developed by Moller [3.12], is actually a family of methods. There are perhaps a dozen or more forms of conjugate gradient al-

gorithms. The methods differ only in their treatment of undetermined systems, accuracy achieved for the problems in hand and their memory requirements etc.

Assuming that we have a quadratic equation that needs to be minimized. The difficulty of using the gradient decent based 1st order algorithms is that, a one dimensional minimization in direction (a) followed by a minimization in direction (b) does not imply that the function is minimized on the subspace generated by (a) and (b). Minimization along direction (b) may in general spoil a previous minimization along direction (a). This is the main reason why a one dimensional minimization in general has to be repeated a number of times, which is sometimes even larger than the number of variables itself. If however the directions are non interfering and linearly independent, at the end of  $N$  steps the process would converge to the minimum of the quadratic equation.

The concept of noninterfering directions is the basis of conjugate gradient method. Two directions  $p_i$  and  $p_j$  are mutually conjugate with respect to a matrix  $G$  if

$$p_i^T G p_j = 0 \quad \text{when } i \neq j \quad (3.12)$$

After minimizing in the direction  $p_i$ , the gradient at the minimizer will be perpendicular to  $p_i$ . If the second minimization is in direction  $p_{i+1}$ , the change of gradient along this new direction say  $g_i$  is represented by :  $g_{i+1} - g_i = \alpha G p_{i+1}$ , where  $\alpha$  is a constant. The matrix  $G$  is the Hessian which contains the second derivatives. Now for the equation (3.12) to be valid, this change is perpendicular to the previous direction [ $p_i^T (g_{i+1} - g_i) = 0$ ], therefore the gradient at the new point remains perpendicular to  $p_i$  and the previous minimization is not spoiled. While, for a quadratic function the conjugate gradient method is guaranteed to converge to the minimum in at most  $(N + 1)$  gradient evaluations, for general functions however it is necessary to iterate until a suitable approximation is obtained.

Let us introduce a vector  $y_k = g_{k+1} - g_k$ . The first search direction  $p_1$  is given by the negative gradient  $-g_1$ . Then the new weight sequence  $w_k$  of the approximations to the minimizer is defined as

$$w_{k+1} = w_k + \alpha_k p_k \quad (3.13)$$

$$p_{k+1} = -g_{k+1} + \beta_k p_k \quad (3.14)$$

where  $g_k$  is the gradient of the error function,  $\alpha_k$  is chosen to minimize the error  $E$  along the search direction  $p_k$  and  $\beta_k$ , which is defined by

$$\beta_k = \frac{y_k^T g_{k+1}}{y_k^T p_k} \quad (3.15)$$

There are many variants of the conjugate gradient methods. The most popular ones being the **Polak - Ribiere** which modified equation (3.15) as

$$\beta_k = \frac{y_k^T g_{k+1}}{g_k^T g_k} \quad (3.16)$$

and the **Fletcher - Reeves** method which modified equation (3.15) as

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad (3.17)$$

Though theoretically, a difficulty with the **Polak - Ribiere** and **Fletcher - Reeves** conjugate methods is that for general functions, the obtained directions are not necessarily the descent directions, however there is no rule which can suggest which of the methods can yield best results for a particular problem. In principal one has to use all the above methods on trial and error basis for the problem in hand.

The backpropagation-momentum algorithm can in principle be considered as an approximate form of conjugate gradient method. In both the cases the gradient direction is modified with a term that takes the previous direction into account, however the parameter  $\beta$  in conjugate gradient is defined by the algorithm itself, while as the momentum rate has to be guessed by the user in case of backpropagation-momentum which may or may not be the optimum value.

### 3.5.2 Lavenberg-Marquardt algorithm

Gradient descent algorithm work well in simple problems but it is too simplistic an approach for real world complex models which can have many free parameters. Convergence therefore can take extremely long time, because of the nature of the gradient descent implementation. For example, when descending the walls of a very steep local minimum bowl we must use a very small step size to avoid ‘rattling out’ of the bowl. On the other hand when we are moving along a gentle sloping part of error surface we should take large steps otherwise it will take forever to reach the minimum. This problem is compounded by the manner in which the gradient descent is implemented. Here, we generally move by taking a step which is some constant times the negative gradient, rather than a step of constant length in the direction of negative gradient. This means that in steep regions (where we have to be careful and not to take large steps) we move quickly and in shallow regions (where we need to take larger steps) we move slowly. The problem is also compounded by the fact that the curvature of the error surface may not always be the same in all directions. For example, if there is a long and narrow valley in the error surface the component of the gradient in the direction that points along the base of the valley is very small while the component perpendicular to the valley walls is quite large, even as, we have to move a long distance along the base and a small distance perpendicular to the walls. Thus one has to use slightly more sophisticated gradient descent algorithms than the simple steepest descent. Using the second order information in other words using the curvature as well as the gradient of the error surface, can speed up the convergence enormously.

## Description of method

A simplistic definition of gradient descent can be

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \mu \mathbf{d} \quad (3.18)$$

where  $\mathbf{d}$  is the average error gradient and  $\mathbf{w}_i$  is the weight matrix.  
Comparing it to the update rule based on quadratic approximation i.e.,

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \mathbf{H}^{-1} \mathbf{d} \quad (3.19)$$

where  $\mathbf{H}$  is the Hessian matrix containing the second derivatives. Quadratic rule implemented here, which is generally the conjugate gradient methods discussed above, is not a universally preferred method since it assumes a linear approximation of functions dependence on  $\mathbf{w}$ , which is however valid only near a minimum. The technique invented by Lavenberg involves "blending" between these two extremes. We can use a steepest descent type method until we approach a minimum, then gradually switch to the quadratic rule. It is possible to guess how close we are to the minimum or how the error is changing. In particular, Lavenberg algorithm is formalized as follows: Let  $\lambda$  be the blending factor which will determine the mix between the steepest descent and the quadratic approximation. The update rule of equation (3.19) above is therefore modified as :

$$\mathbf{w}_{i+1} = \mathbf{w}_i - (\mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{d} \quad (3.20)$$

where  $\mathbf{I}$  is the identity matrix. As  $\lambda$  gets small, the rule approaches the quadratic approximation update rule above. If  $\lambda$  is large, the rule approaches

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \frac{1}{\lambda} \mathbf{d} \quad (3.21)$$

which is the steepest descent. The algorithm adjusts  $\lambda$  according to whether the error function  $\mathbf{E}$  is increasing or decreasing. This is done as follows:

1. Do an update as directed by the rule above
2. Evaluate the error as the new weight vector
3. If the error has increased as a result of update, then retract the step (i.e, reset the weights to their previous values) and increase  $\lambda$  by a factor  $\approx 10$  or some such significant factor. Then go to (1) and try the update again.
4. If the error has decreased as a result of update then accept the step (i.e keep the weights at their new values) and decrease  $\lambda$  by a factor  $\approx 10$  or so.

The intuition is that if the error is increasing, our quadratic approximation is not working well and we are likely not near a minimum, so we should increase  $\lambda$  in order to blend more towards simple gradient descent. Conversely, if error is decreasing, our approximation is working well, and that we are getting closer to minimum, so  $\lambda$  is decreased to bank more on Hessian.

Marquardt improved this method with a clever incorporation of estimated local curvature information, resulting in the Levenberg-Marquardt method [3.13]. The insight of Marquardt was that when  $\lambda$  is high and we are doing essentially gradient descent, we can still get benefit from the Hessian matrix that we estimated. In essence Marquardt suggested that we should move further in the direction in which the gradient is smaller in order to get around the classic "error valley" problem. So Marquardt replaced the identity matrix in Levenberg's original equations with diagonal of the Hessian.

$$\mathbf{w}_{i+1} = \mathbf{w}_i - (\mathbf{H} + \lambda \text{diag}[\mathbf{H}])^{-1} \mathbf{d} \quad (3.22)$$

It is important to know that the method is nothing more than a heuristic method. It is not optimal for any well defined criterion of speed or final error, but is merely a well thought out optimization procedure. However, it is one method that works extremely well in practise, though there is a little understanding for the reason behind this. Lavenberg-Marquardt algorithm has presently become a virtual standard for optimization of non-linear models. It's only flaw seems to be that it requires a matrix inversion step as part of the update scheme which scales as the  $N^3$  where  $N$  is the number of weights. For medium sized networks this method may not be too slow, however for larger networks the cost of matrix inversion kills the advantage gained by cleverness of the method as it is lost in time taken to do each iteration [3.14].

### 3.5.3 Radial basis functions

Radial basis functions (RBF) neural networks have recently been studied intensively. In many applications, several results have been obtained using this method. The RBF neural network has the universal approximation ability, therefore, the RBF neural network can be used for the interpolation problem [3.15]. The Radial basis function networks (RBFNs), also called the potential function networks, have been studied as an alternative to the multilayered feedforward neural networks (MFNNs). A RBFN is a multidimensional non linear function mapping that depends on the distance between the input vector and the center vector. A RBFN with  $n$ -dimensional inputs and a single output ( $y$ ) can be represented by the weighted summation of finite number of radial basis functions as follows:

$$y = f(\mathbf{x}) = \sum_{i=1}^n w_i \phi_i(||\mathbf{x} - \mathbf{c}_i||) \quad (3.23)$$

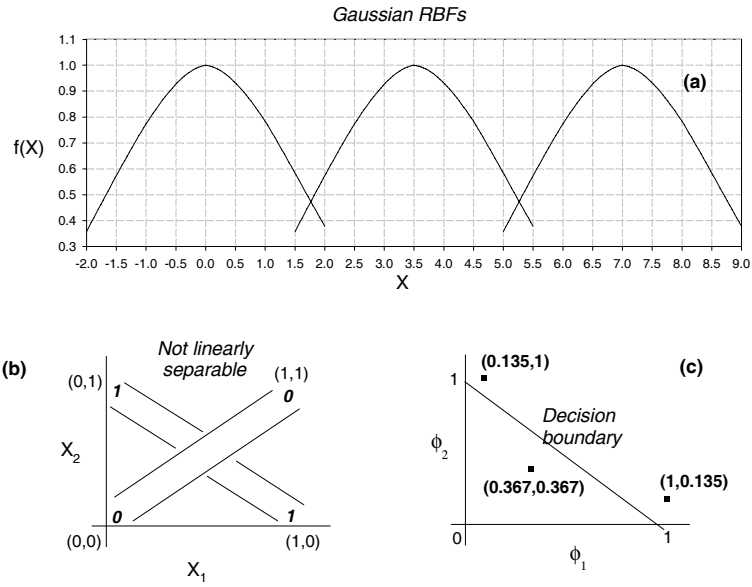


Figure 3.7: (a) Combination of Radial Basis Function Neural Networks (b) XOR which is a not a linearly separable problem. (c) Using RBF a linearly inseparable problem becomes separable.

where  $\phi_i(||\mathbf{x} - \mathbf{c}_i||)$  is the radial basis function of  $\mathbf{x}$ ,  $\phi(\cdot)$  is an arbitrary non linear function,  $||\cdot||$  denotes a norm that is usually assumed to be Euclidean, the known vectors and  $\mathbf{c}_i$  are viewed as the centers of radial basis functions. Fig 3.7a shows a combination of 3 gaussian RBF's with basis functions located at 0, 3.5 and 7 respectively. Gaussian function is the most commonly used RBF. The application of RBF in neural networks can be seen by revisiting the conventional XOR table shown in Table 3.1.

Table 3.1: Truth table for an XOR non linear problem.

$X_1$	$X_2$	output
0	0	0
1	0	1
0	1	1
1	1	0

The output of the truth table, represented in the last column, is a linearly inseparable problem as shown in Fig 3.7b and no single line (decision boundary) can be drawn to separate these. In order to find a way to separate these two classes represented by  $x_1, x_2$ , let there be a transformation from the input space to some



arbitrary ( $\phi$ ) space denoted by  $(\phi_1, \phi_2)$ , and represented in Fig 3.7c above. Let the transformation axis be represented by a gaussian RBF of the form

$$\phi_1(\mathbf{x}) = \exp(-(\|\mathbf{x} - \mathbf{t}_1\|)^2) \quad (3.24)$$

where  $\mathbf{t}_1$  represents the center of the function. Let this be assumed to be represented numerically by  $\mathbf{t}_1=[1,1]$ . Similarly let the other axis be represented as:

$$\phi_2(\mathbf{x}) = \exp(-(\|\mathbf{x} - \mathbf{t}_2\|)^2) \quad (3.25)$$

where like  $\mathbf{t}_1$ ,  $\mathbf{t}_2$  be represented as  $\mathbf{t}_2=[0,0]$ . Simple center values of  $\mathbf{t}_1$  and  $\mathbf{t}_2$  have been chosen for convenience. Now let us evaluate the input space (X) in terms of the transformed space  $\phi$ . The transformation will have to be computed for all the values of the input space represented by the truth table above. Computation is shown for a specific case when  $X_1=1$ ,  $X_2=1$  from the truth table [3.1] and  $t_1 = [1,1]$ . Thus replacing the values in equation (3.24) we have :

$$\phi_1(\mathbf{x}) = \begin{cases} \exp\{ - [(\mathbf{x}_1 - 1)^2 + (\mathbf{x}_2 - 1)^2] \} \\ \exp\{ - [(1 - 1)^2 + (1 - 1)^2] \} \\ \exp\{ -0 \} \\ 1 \end{cases} \quad (3.26)$$

Similarly for equation (3.25) we have for  $\phi_2$  :

$$\phi_2(\mathbf{x}) = \begin{cases} \exp\{ - [(\mathbf{x}_1 - 0)^2 + (\mathbf{x}_2 - 0)^2] \} \\ \exp\{ - [(1 - 0)^2 + (1 - 0)^2] \} \\ \exp\{ -2 \} \\ 0.135 \end{cases} \quad (3.27)$$

After evaluating the other values the new truth table is represented in Table 3.2.

Table 3.2: Truth table after transformation from input space to the  $\phi$  space.

$X$	$\phi_1$	$\phi_2$
1, 1	1	0.135
0, 1	0.367	0.367
0, 0	0.135	1
1, 0	0.367	0.367

The plot of the truth table mentioned above is shown in Fig 3.7c. It is clear from the figure that after the RBF transformation, a linearly inseparable problem is converted to a linearly separable problem with a clean decision boundary as represented (Fig 3.7c). Thus patterns which were not separable in the input space (X) are now separated in the new space ( $\phi$ ).

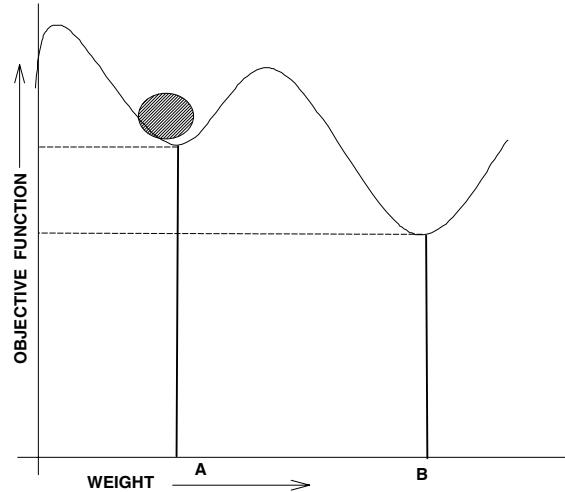


Figure 3.8: Local minima problem demonstrated for Simulated annealing method.

### 3.5.4 Simulated annealing

Simulated annealing is a generic probabilistic algorithm for global optimization problem, namely locating a good approximation to the global optimum of a given function in a large search space [3.16]. Simulated annealing algorithm implemented in ANN is done by making certain pseudorandom changes that result in improvements. The training process, tends to minimise the objective function that can otherwise get trapped in a poor solution. Fig 3.8 demonstrates how this happens in a system with say a single weight.

Assume the weight is set initially to the value at point  $A$ . If the random weight steps are small, all deviations from point  $A$  increase the objective function (error function) and hence will be rejected. The superior weight setting at point  $B$  will never be found and the system will be trapped in 'local minimum' instead of the 'global minimum' at point  $B$ . If the random weight adjustments are very large, both point  $A$  and  $B$  will be visited frequently, but so will be every other point. The weight will change so drastically that it will never settle into the desired minimum at  $B$ .

A useful strategy to avoid this problem is to start with large steps and then gradually reduce the size of average random step. This allows the network to escape local minima, while ensuring the network stabilization.

Local minimum entrapment plagues all 'minimum-seeking' training algorithms. This includes the standard backpropagation and other 1st order networks as well and represents a serious and widespread difficulty that is often overlooked. Statistical methods can overcome this problem by a weight adjustment strategy that causes the weights to assume the globally optimal value which can eventually make reaching the point  $B$  possible.

As an explanatory analogy, suppose the figure represents a marble on a surface in a box. If the box is shaken violently in the horizontal direction, the marble will move rapidly from side to side. Never settling at one point, at any instant the marble maybe at any point on the surface with equal probability.

If the violence of the shaking is gradually reduced a condition will be reached when the marble sticks briefly at point  $A$  and point  $B$ . At a still lower level of shaking, the marble will stay at both points  $A$  and  $B$  for short times. If the shaking is continually reduced, a critical point will be reached where shaking is just strong enough to move marble from point  $A$  to  $B$ , but not strong enough to enable the marble to climb the hill from  $B$  to  $A$ . Thus, the marble will end up in global minimum as the shaking amplitude is reduced further and eventually to zero.

ANN's are trained in essentially the same way through random adjustments of weights. At first, large random adjustments are made, retaining only those weight changes that reduce the objective function. The average step size is then gradually reduced and the global minimum will eventually be reached.

The procedure has a strong resemblance to the annealing of metals; hence the term 'simulated annealing' is used to describe this mechanism. If the metal is raised to a temperature above its melting point, the atoms are in violent random motion. As with all physical systems, the atoms tend to settle towards minimum energy state, but at high temperatures the vigour of the atomic motions prevents this. As the metal is gradually cooled, lower and lower energy states are assumed until finally the lowest of all possible states, a global minimum, is achieved. In annealing process, the distribution of energy states is determined by the relationship:

$$P(e) \propto \exp(-e/kT)$$

where  $P(e)$  is the probability that the system is in a state with energy  $e$ ,  $k$  is the Boltzman's constant and  $T$  is temperature in degrees Kelvin. At high temperatures  $P(e)$  approaches one for all energy states. Thus a high energy state is almost as likely as a low energy state. As the temperature is reduced, the probability of high energy states decreases as compared to the probability of low energy states. As the temperature approaches zero it becomes very unlikely that the system will exist in a high energy state.

## Training

Applying this statistical method is quite straight forward

1. Define a variable  $T$  that represents an artificial temperature. Start with a large value of  $T$ .
2. Apply a set of inputs to the network, and calculate the outputs and objective function.

3. Make a random weight change and re-calculate the network output and the change in objective function due to the weight change.
4. If the objective function is reduced (improved) retain the weight change.

If the weight change results in an increase in objective function, calculate the probability of accepting the change from the Boltzmann distribution as follows

$$P(c) \propto \exp(-c/kT)$$

where  $P(c)$  is the probability of a change of  $c$  in the objective function,  $k$  is a constant analogous to the Boltzmann constant that must be chosen for the problem at hand, and  $T$  is the artificial temperature. Select a random number ' $r$ ' from a uniform distribution between zero and one. If  $P(c)$  is greater than ' $r$ ', retain the change, otherwise return the weight to previous value.

This allows the system to take an occasional step in the direction that worsens the objective function, thereby permitting it to escape a local minimum where any small step raises the objective function. To complete the Boltzmann training strategy, repeat the steps 3 and 4 above, over each of the weights of the network, gradually reducing the temperature  $T$  until an acceptably low value for the objective function is achieved. At this point, a different input vector is applied and the training process is repeated. The network is trained on all vectors in the training set, perhaps repeatedly until the objective function value is acceptable for all.

The size of the random weight change in step 3 can be determined in many ways, e.g., emulating the thermal system, the weight change  $w$  can be selected according to the Gaussian distribution.

$$P(w) = \exp(-w^2/T^2) \quad (3.28)$$

where  $P(w)$  is the probability of a weight change of size  $w$  and  $T$  is artificial temperature. This weight change selection method produces a system analogous to the method described in [3.17].

### 3.5.5 Secant Methods-One step secant

In the above second order methods, we have seen that the Hessian matrix, plays a crucial role in minimising the error function. If however the Hessian matrix is not available analytically, secant methods (traditionally referred to as the quasi-Newton methods) are the widely used methods for approximating the Hessian in an iterative way using only information about the gradient. In one dimension the second derivative  $\frac{\partial^2 E(w)}{\partial w^2}$  can be approximated with the slope of the secant through values of the first derivative in two near points.

$$\frac{\partial^2 E(w)}{\partial w^2}(w_2 - w_1) \approx \left[ \frac{\partial E(w_2)}{\partial w} - \frac{\partial E(w_1)}{\partial w} \right] \quad (3.29)$$

The one step method is an approximation of the Gauss Newton method for error minimization. The advantage of this method is smaller memory requirements and lesser computation time[3.18]. One step secant method has been an extensively used ANN model.

## 3.6 Hybrid algorithms

### 3.6.1 Higher order neurons

The Standard Neuron is a combination of aggregation and activation functions. The standard form of aggregation function is generally the linear weighted sum (linear basis function). The most common activation functions as mentioned previously are the sigmoid and the tangent hyperbolic functions. Such neural models, however, when used to solve real life problems may require a large number of neurons in the standard neural network. Also the number of unknowns to be determined grows with the number of neurons and the hidden layers. Therefore, the working of ANN becomes computation and memory intensive. The computational burden can be reduced either by reducing the number of neurons or by improving the learning techniques. The number of neurons in the ANN in turn depends on the neuron model itself. The standard neuron model is a first order model. This has limitations of mapping and classification. These limitations can be overcome either by using second order gradient techniques discussed earlier or by using higher order neuron models. The application of second order gradient techniques, such as conjugate gradient and quasi-Newton methods, instead of simple gradient technique, have shown to achieve rapid convergence. Second order gradient techniques may reduce the architectural complexity but not the learning complexity.

As the mapping and the classification power of a neuron depends on its order, a higher neuron model is likely to possess better mapping and classification capability. Higher order neuron model, which includes quadratic and higher order basis functions in addition to linear basis function reduces the learning complexity. The architectural complexity increases with the number of higher order basis functions. The overall complexity of working of the ANN with higher order neurons is less compared to second order gradient techniques.

#### Implementation

A higher order neuron model is considered which has  $N$  aggregation functions and  $F$  activation functions [3.19]. The aggregation functions can be linear weighted sum (linear basis function), quadratic or the higher order basis functions. The activation functions can be a linear or preferably some non-linear function.

The cross-products of the input terms is added into the model, where each component of the input pattern multiplies the entire input pattern vector. A reasonable

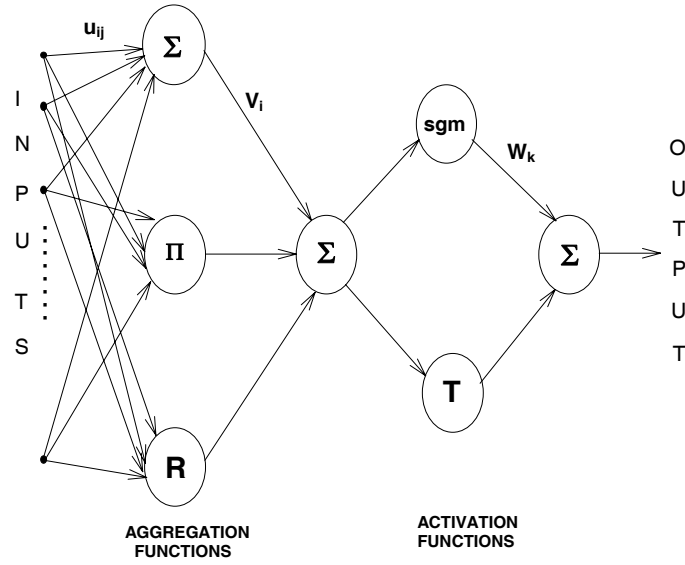


Figure 3.9: A higher order Neuron model with 3 aggregation functions and 2 activation functions.

way to do this is to add all interaction terms between input values. For example, for a back-propagation network with three inputs (A, B and C), the cross-products would include: AA, BB, CC, AB, AC, and BC, which adds second-order terms to the input structure of the network.

Additional input nodes can be a functional expansion of the base inputs. Thus, a backpropagation model with A, B and C might be transformed into a higher-order neural network model with inputs: A, B, C,  $f(A,B,C)$ ,  $g_1(A,B)$ ,  $g_2(B,C)$  etc. In this model, input variables are individually acted upon by appropriate functions. Many different functions can be used. The overall effect is to provide the network with an enhanced representation of the input. No new information is added, but the representation of the inputs is enhanced. Higher-order representation of the input data can make the network easier to train. The joint or functional activations become directly available to the model. Higher order neuron with three aggregations and two activations is shown in Fig 3.9.

Here,  $\Sigma$  represents the weighted sum of inputs,  $\Pi$  represents sum of products of weighted inputs and 'R' represents the radial basis term as discussed above. *Sgm* and *T* represent the sigmoidal activation and the Tan hyperbolic functions respectively.

In some cases, a hidden layer is no longer needed. However, there are also certain limitations to this network model. More input nodes must be processed to use the transformations of the original inputs. With higher-order systems, the problem

can sometimes get worsened and not help at all. Yet, because of the limitations imposed on finite processing time in some cases, it is important that the inputs are not expanded more than what is needed to get an accurate solution.

### 3.6.2 Neuro Fuzzy systems

Several other attempts apart from the usage of ANN have been reported to understand and model the capabilities of the human brain. One such popular technique is the use of Fuzzy Logic algorithms [3.20]. This algorithm represents an altogether different aspect of human information processing. However, the model presented may not perform all functions performed by the brain independently. Every intelligent technique has particular computational properties (e.g. ability to learn, explanation of decisions) that make them more or less suited for a particular set of problems. While neural networks are good at recognizing the patterns, they are not good at explaining how they reach their decisions. Fuzzy logic systems, which can reason with imprecise information, are good at explaining their decisions but they, unlike the neural networks systems, can not automatically acquire the rules they use to make those decisions. These limitations have been a central driving force behind the creation of intelligent hybrid systems where two or more techniques are combined in a manner that overcomes the limitation of individual techniques. The computational power of Neuro Fuzzy systems, which are fuzzy rule based systems, are implemented in the framework of neural networks.

The behaviour of fuzzy systems is designed simply by a fuzzy rule :

IF < premise > THEN < consequent >

which uses linguistic variables with symbolic terms, each term representing a fuzzy set. The fuzzy inference mechanism consists of three stages: in the first stage, the values of the numerical inputs are mapped by a function according to a degree of compatibility of the respective fuzzy sets, this operation can be called **fuzzyfication**. In the second stage, the fuzzy system processes the rules in accordance with the firing strengths of the inputs. In the third stage, the resultant fuzzy values are transformed again into numerical values, this operation can be called **defuzzyfication**. Essentially, this procedure makes possible the use of fuzzy categories in representation of words and abstract ideas of the human beings in the description of the decision taking procedure.

There are three main types of Neuro-fuzzy algorithms [3.21].

1. OR/AND Neuron based Neural Network (FNN)
2. Adaptive Neuro Fuzzy Inference Systems (ANFIS)
3. Compensatory Neuro Fuzzy Systems (CNFIS)

**AND and OR Neurons**

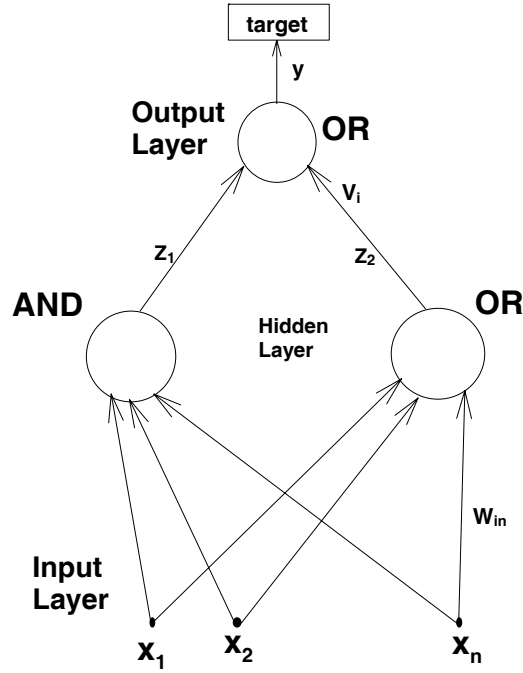


Figure 3.10: Standard architecture of an AND/OR Neuron consisting of 3 layers.

The AND neuron aggregates input signals  $x = [x_1, x_2, \dots, x_n]$  by combining them individually with the connections weights  $w = [w_1, w_2, \dots, w_n]$  and globally ANDing these results,

$$y = AND(x, w)$$

The structure of the OR neuron is dual to that reported for the AND neuron, namely,

$$y = OR(x, w)$$

The AND and OR neurons realize 'pure' logic operations on the inputs. The implemented OR/AND neuron constitutes a three layer network and is constructed by arranging the above mentioned AND and OR neurons into a structure displayed in Fig 3.10. Due to the strong functional cohesion existing between the neurons, it is called OR/AND neuron. The pure AND and OR characteristics produced at the level of the hidden layer are then combined by the OR neuron constituting the output layer. Changing the connection weights between hidden and output layer, an entire range of intermediate logical characteristics of the structure spread between the AND-like and OR-like functional behavior can be obtained.

## Learning



Learning in the above OR/AND neuron was addressed as a problem of supervised training which concerns a series of modifications of the connections (parametric learning) using gradient descent algorithm. The successive updates of the connections  $W = [w_{ij}]$  and  $V = [v_1, v_2]$  (weights connecting the AND/OR layer to output) within this training are controlled by the gradient of the predetermined performance index, in our case the mean squared error performance criterion (Q). Considering the OR/AND neuron without the nonlinear element, we have from Fig 3.10,  $y = \text{OR} < z_1, z_2 > V$ ;  $z_1 = \text{AND}(x, w_1)$ ;  $z_2 = \text{OR}(x, w_2)$ . Thus at the two layers we have:

$$\begin{aligned}\frac{\partial Q}{\partial w_{ij}} &= -2(\text{target} - y) \frac{\partial y}{\partial z_i} \frac{\partial z_i}{\partial w_{ij}} \\ \frac{\partial Q}{\partial v_i} &= -2(\text{target} - y) \frac{\partial y}{\partial v_i}\end{aligned}\quad (3.30)$$

where  $i = 1, 2, \dots, n$  and  $Q = (\text{target} - y)^2$ .

The nonlinear sigmoidal element placed in series with the OR neuron of the output layer is given by,

$$OUT = \frac{1}{1 + \exp[-(y - m)\sigma]} \quad (3.31)$$

Both the parameters of the nonlinearity ( $m$  and  $\sigma$ ) were subjected to changes during the learning of the neuron. The learning formulae become slightly modified as compared to what has been derived above, and take the form:

$$\begin{aligned}\frac{\partial Q}{\partial w_{ij}} &= -2(\text{target} - y) y (1 - y) \sigma \frac{\partial y}{\partial z_i} \frac{\partial z_i}{\partial w_{ij}} \\ \frac{\partial Q}{\partial v_i} &= -2(\text{target} - y) y (1 - y) \sigma \frac{\partial y}{\partial v_i} \\ \frac{\partial Q}{\partial m} &= -2(\text{target} - y) y (1 - y) (-\sigma) \\ \frac{\partial Q}{\partial \sigma} &= -2(\text{target} - y) y (1 - y) (y - m)\end{aligned}\quad (3.32)$$

### Disadvantages of the OR/AND Neural Network

Although the number of neurons and hidden layers required are less than that required for standard Neural Network, the reduction in error is not always assured.

### 3.6.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

An adaptive network, as its name implies, is a network structure consisting of nodes and directional links through which the nodes are connected. Moreover, part or all

of the nodes are adaptive, which means their outputs depend on the parameters pertaining to these nodes. In an adaptive network, each node performs a particular function on incoming signals as well as a set of parameters pertaining to this node. The node can be either a square or a circle node. A square node has some inherent parameters while a circle node does not have any.

### ANFIS architecture and learning

In ANFIS, the input space is partitioned according to the number of membership functions selected for each of the inputs. A membership function provides a measure of the degree of similarity of elements to a fuzzy set. The number of rules ( $nr$ ) is given by the number of membership functions for each input ( $nmf$ ) raised to the power of the number of inputs ( $ni$ ). i.e.,

$$nr = (nmf)^{ni}$$

After the number of membership functions for each of the inputs is known, the input space is partitioned accordingly and one of the rules as discussed, will be used to approximate the function. In order to modify the parameters associated with the rules adaptive network concepts are utilized. For simplicity, consider a fuzzy inference system with two inputs  $x$  and  $y$  and one output  $z$ , assuming the rule-base contains two types of fuzzy rules which are:

- Rule 1 : If  $x$  is  $A_1$  and  $y$  is  $B_1$  , then  $f_1 = p_1x + q_1y + r_1$
- Rule 2 : If  $x$  is  $A_2$  and  $y$  is  $B_2$  , then  $f_2 = p_2x + q_2y + r_2$

where  $p, q, r$  subscripts are referred to as the node parameters. The corresponding architecture is shown in Fig 3.11

The node function of the different layers is given below.

- Layer - I:

$$O_i^1(x) = \frac{1}{1 + \left[ \left( \frac{x-c_i}{a_i} \right)^2 b_i \right]} \quad (3.33)$$

Where  $x$  is the input to node  $i$  (Layer-I) and  $a_i, b_i, c_i$  is the node parameter set. Parameters in this node are referred to as premise parameters.

- Layer - II:

Every node in this layer is a circle node, which multiplies the incoming signals and sends the product out.

$$w_i = O_i^1(x) \times O_j^1(y) \quad (3.34)$$

Each node output represents the firing strength of a rule.

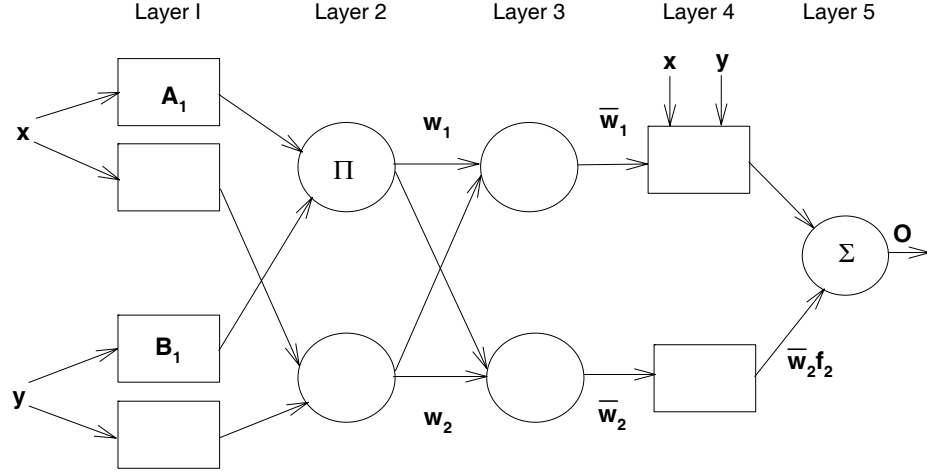


Figure 3.11: Figure shows the Fuzzy architecture

- Layer - III:  
The nodes in this layer normalize the outputs of all layer 2 nodes.

$$\overline{w}_i = \frac{w_i}{\sum_i w_i} \quad (3.35)$$

Outputs of this layer are called normalized firing strengths.

- Layer - IV:  
Every node in this layer is a square node with a node function

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \quad (3.36)$$

Parameters in this layer are referred to as consequent parameters.

- Layer-V:  
This layer has the single node that computes the overall output

$$O_i^5 = \sum_i \overline{w}_i f_i \quad (3.37)$$

After the node output at the output layer above is computed, error is minimized using the gradient descent method which is essentially the same treatment followed by standard backpropagation method.

### 3.6.4 Compensatory Neuro-Fuzzy system

The selection of optimal fuzzy operations and optimal reasoning mechanism is carried out by using a compensatory parameter in fuzzy logic, which can be updated during learning process so that it automatically picks up the degree of compensation required between the pessimistic (min.) and optimistic (max.) operations. The neurofuzzy system incorporated with compensatory reasoning method is known as the compensatory neuro-fuzzy system.

#### Architecture

The Compensatory Neuro-Fuzzy System can have five the following five functional layers: They are called as :

- input layer
- fuzzification layer
- pessimistic-optimistic layer
- compensatory operation layer
- defuzzification layer.

A compensatory neuro-fuzzy system is initially constructed layer by layer according to linguistic variables, fuzzy IF-THEN rules, the pessimistic and optimistic operations, the fuzzy reasoning method, and the defuzzification scheme of a fuzzy logic control system. For clarity and convenience, we can define more extensive fuzzy neurons which perform some kinds of fuzzy operations. Since the fuzzy logic control system performs typical functions of fuzzification, fuzzy reasoning, and defuzzification, the control-oriented fuzzy neurons can be defined as the fuzzy neurons performing these functions. Generally, a fuzzy neuron which performs fuzzification is called a fuzzification neuron, a fuzzy neuron which performs fuzzy reasoning is called a fuzzy-reasoning neuron and a fuzzy neuron which performs defuzzification is called a defuzzification neuron[3.22].

### 3.6.5 Fuzzification and decision oriented fuzzy neuron

A typical fuzzification neuron is a fuzzy linguistic term of the fuzzy linguistic variable . The fuzzification neuron performs a mapping from a crisp value 'x' into a fuzzy set,  $A_i$  with a degree 'y' such that  $y = \mu A_i(x)$  where  $\mu$  is the membership function.

- **Fuzzy-reasoning neuron** A simple fuzzy-reasoning neuron, can map the inputs  $x_i$  ( $i=1,2,...,n$ ) to the output 'y' to the output through some t-norm function  $T(x_1, x_2, ..., x_n)$  e.g,

$$T(x_1, x_2, ..., x_n) = \text{Min}(x_1, x_2, ..., x_n)$$

$$T(x_1, x_2, ..., x_n) = (x_1, x_2, ..., x_n)$$

The complex fuzzy-reasoning neuron can be constructed by using T-norm fuzzy neurons performing If-condition matching of fuzzy logic rules and a T-conorm fuzzy neuron integrating the fired rules.

- **Defuzzification neuron**

A typical defuzzification neuron can generate the final crisp value 'y' based on inputs ' $x_i$ ' ( $i=1,2,...,n$ ) and the weights  $w_i$  ( $i=1,2,...,n$ ). Here, the weights  $w_i$  are the parameters of the output membership functions. A detailed defuzzification scheme is described in [3.23].

- **Decision-oriented fuzzy neurons**

Depending on their operations the Compensatory neurons are defined as pessimistic or optimistic neurons

- **Pessimistic fuzzy neurons**

The pessimistic fuzzy neuron can map the inputs  $x_i$  ( $i=1,2,...,n$ ) to a pessimistic output by making a conservative decision for the situation by considering the worst case scenario. T-norm fuzzy neurons are pessimistic neurons.

- **Optimistic fuzzy neurons**

On the other hand, the optimistic fuzzy neuron, can map the inputs  $x_i$  ( $i=1,2,...,n$ ) to the optimistic output by making an optimistic decision for the situation by considering the best case scenario. Tco-norm fuzzy neurons are optimistic neurons.

- **Compensatory fuzzy neurons**

The compensatory fuzzy neuron can map the pessimistic input  $x_1$  and the optimistic input  $x_2$  to the compensatory output  $y$  to make a relatively compromised decision for the situation between the worst case and the best case.

### 3.7 Techniques to improve results in ANN

Fast training is desirable, but ultimately one wants the best possible results on the test set. This section describes the techniques for getting good performance of the neural network models. The following points will focus on good learning.

### 3.7.1 Overfitting, generalization and size of networks

One of the virtues of ANN is that it will fit any function even if the form of the function is not known. This however, is a shortcoming also sometimes as it can lead to over-fitting, which means as training goes on, the network will end up fitting the training data set very closely while ruining the results on the test set. Thus we must be careful to train the network until the test set results hit a minimum and then stop. However, one does not always know when the test set has hit a minimum so as to stop training, more so a local minimum may also hit this process. Over-fitting generally occurs when we aim for rigorous error levels with a very large architecture with the given training set. The net, once trained, gives excellent results with the training set but totally adverse results with testing set. The error goal and network size needs to be judiciously decided depending on the nature of the problem, number of training data available, closeness of training data etc.

- **Generalization**

In classification problems, there is no guarantee that the trained neural network will suggest a sensible way to partition the boundaries between classes. There are chances that the testing data may be highly misclassified even though the training data gets classified properly. Odd generalisations can be minimised somewhat by averaging results over a number of networks. A reasonable cross check is to do a nearest neighbourhood analysis of each test pattern as well.

- **Size of the network**

There is unfortunately no established method to decide on the size with respect to the number of hidden nodes that can be used. Literature is full of different methods (some of them even contradictory) which can be employed for evaluating the number of hidden nodes needed for the problem at hand [3.24-3.27]. A thumb rule, which some neural network expert's use, is that the number of hidden neurons required is two-third of the sum of input neurons and output neurons. Another thumb rule is that we need at least as many training set patterns, as there are weights in the network. No firm rule can therefore be stated regarding the size of the network - one has to evaluate these through the trial and error method or do rigorous experimentation with the neural model. An easier way that we have employed is the use of Singular Vector Decomposition for finding the optimum nodes in hidden layer. The details of this method will be discussed in Chapter 5.

### 3.7.2 Combining network outputs and pruning

One of the techniques used to improve results is to combine estimators (network outputs) of many networks, the equivalent of getting the opinion of many experts.

One of the notable advantages of this method is that one may get very good results even with rather poorly trained networks.

- **Pruning of the network**

Pruning a network is the process of removing unnecessary processing elements and connections (weights). It is often desirable to prune a network before deploying it for operation to reduce the I/O processing time. Pruning can often be performed without any noticeable effect to the performance of a network. It can be performed by examining the output of units across the training set inputs and eliminating non-contributory units and weights. Pruning of the network however can also be avoided using the Singular Vector Decomposition methodology.

---

## 3.8 References

- [3.1] R. Beale, T. Jackson, An Introduction to Neural Computing, Adan Higler, Bristol, England (1990).
- [3.2] I. Aleksander, Impossible Minds: My neurons, My Consciousness published by Imperial College Press (1996).
- [3.3] L. Fausett, Fundamentals of Neural Networks, architectures, algorithms and applications (1994).
- [3.4] R. Harvey, Neural Network Principles, Prentice Hall Publishing (1994).
- [3.5] W. S. McCulloh and W. Pitts, Bulletin of Mathematical Biophysics, 5 (1943) 115.
- [3.6] T. Kohonen, Complex Systems, 2 (1988) 321.
- [3.7] G.A. Carpenter, S. Grossberga, Comp. Vision, Graphics and Image Processing 37, (1987) 54.
- [3.8] J. Geobel et al, Lett Editor, Astron Astrophys 222, (1989) L5.
- [3.9] D. E. Rumelhart et al., Nature, 323 (1986) 533.
- [3.10] M. Reidmiller, Computer Standards Interfaces, 16 (1994) 265.
- [3.11] S. E. Fahlman, Proc. Connectionist Models Summer School, Morgan-Kaufmann, Los Altos CA. CMU-CS-88-162 (1988).
- [3.12] M.F. Moller, Neural networks 6 (1993) 525.

- [3.13] W. H. Press et al., Numerical Recipies in C++, 668 (2002).
- [3.14] D. W. Marquardt, J Soc. Indus. Appl. Math 11 (1963) 431.
- [3.15] J. Park et al., Neural Computation 3 (1991) 246.
- [3.16] S. Kirkpatrick et al., Science 220 (1983) 671.
- [3.17] N. Metropolis et al., Jour of Chemical Phys, 21 (1953) 1087.
- [3.18] R. Battiti et al., Neural Computation, 4 (1992) 141.
- [3.19] C. T. Giles et al., Appl. Opt 26 (1987) 4972.
- [3.20] L. A. Zadeh, Information and Control 8 (1965) 338.
- [3.21] C. T. Lin et al., Neural-Fuzzy systems, Prentice Hall, USA, (1996).
- [3.22] Yan-Qing Zhang et al., IEEE trans. on Neural Networks, 9 (1998) 83.
- [3.23] D. C. Kuncicky et al., Proc. 3rd IFSA Congr. (1989) 113.
- [3.24] G. Mirchandani, IEEE Trans. on Sys. and Circuits, 36 (1989) 661.
- [3.25] L. Fletcher et al., IEEE Trans. Neural Networks 1608 (1998).
- [3.26] Shuxiang Xu et al., 5th Intl Conf IT and Appl., ICITA (2008) 683.
- [3.27] B. Choi et al., Neurocomputing 71 (2008) 3640.



# Chapter 4

## Comparative study of ANN algorithms

### 4.1 Introduction

An important research activity in the field of neural networks is to compare the performance of different ANN algorithms on benchmark problems so as to develop more efficient algorithms for solving real world problems which contain noisy and scarce data. It has also been observed by several workers, that neural network algorithms are often benchmarked rather poorly [4.1].

The aim to construct a neural network training benchmark arises from the situation that ANNs are very powerful information processing tools but their functionality is rather poorly understood. It has been postulated that a properly chosen ANN model can approximate any continuous function to an arbitrary accuracy [4.2-4.3]. However it is also evident that neural networks are difficult to train in practise. Extensive training times and the concern towards the improvement have led to a rich set of proposed alternatives for a variety of network construction activities. As mentioned in the previous chapter, these activities include a variety of activation functions, network architectures, error minimization algorithms, node optimization methods, network pruning methods, etc. With a large number of combinations of these alternatives and the likelihood that the training method performance is problem dependant, it has become increasingly difficult to establish any useful guidelines for training of the ANNs.

It is thus a belief that benchmarking the ANNs, that is, a comprehensive set of training problems and evaluation criteria, would be helpful in a number of ways. Firstly, it would provide a useful framework for understanding old algorithms and development of the newer ones. A comprehensive set of training problems would assist algorithm development in terms of understanding the strength/weakness of the method. A set of evaluation criteria can ensure proper weightage to the factors such as algorithm sensitivity parameters, in addition to the parameters like exe-

cution times etc. Secondly, a proper benchmark problem can encourage controlled experiments to better understand the dynamics of neural network training. Thirdly, benchmarking could assist in comparison of the algorithms and provide guidance if a particular algorithm is better compared to others and under what conditions.

There have been a number of standard problems used for evaluating training algorithms, ranging from the XOR problem, to the spiral problem for the classification. More importantly, it has also been observed in the literature that performance of any algorithm is only compared to the standard backpropagation algorithm alone [4.4] even though there are several powerful and widely used algorithms readily available now. Keeping this in mind, we carried out a detailed study where the performance of three generations of neural network algorithms i.e 1st order algorithms (Standard Backpropagation and Resilient Backpropagation), 2nd order algorithms (Conjugate Gradient, Levenberg-Marquardt, Radial Basis Function, Simulated Annealing), the Hybrid models like the Higher Order Neuron model and Neuro-Fuzzy system, is examined by applying them to standard benchmarking problems like IRIS data, XOR/N-Bit parity and Two Spiral data. In addition to benchmark problems discussed above, we have also applied the above mentioned neural network algorithms for solving several regression problems such as  $\cos(x)$  and a few special functions like the Gamma function, the complimentary Error function and the upper tail cumulative  $\chi^2$ -distribution function.

## 4.2 Benchmarking of ANN algorithms

The comparative performance of the ANN algorithms described above have been studied by applying them to standard benchmarking problems like IRIS data, XOR/N-Bit parity, Two Spiral data and Cosine(x). While we used standard ANN package contained in the MATLAB software for implementing the Lavenberg-Marquardt algorithm, the implementation of other algorithms like Backpropagation, Resilient Backpropagation, Conjugate Gradient Method, Radial Basis Functions, Simulated Annealing, Neuro-Fuzzy etc., were done by using the dedicated ANN simulator package BIKAS (BARC - IIT Kanpur ANN Simulator)[4.5]. Written in Java environment, this dedicated ANN package contains a variety of neural network algorithms like the standard backpropagation, resilient, scale and self conjugate, higher order network functions, simulated annealing and radial basis methods, adaptive resonance theory algorithms, self growing networks and fuzzy algorithms. An exhaustive library of about 15 error minimization functions (like the conventional RMS error function, Hyperbolic square error, Minkowski error, Hubers error function, Cauchy error function etc.) and about 25 activation functions (like the sigmoid function, hyperbolic tan, sine, cosine, decaying exponential, Gaussian, bipolar logarithmic etc.) are also provided in this package.

The training and testing of all the ANN algorithms used in the present work has been done on a Pentium P-III 700 MHz machine. Rigorous checks were also

performed at various stages to ensure that the ANN configuration used for a particular problem was properly optimized with respect to the number of nodes in the hidden layer. This was done by monitoring the RMS error while training the ANN. The RMS error employed here is defined as :

$$RMS = \frac{1}{PI} \sqrt{\sum_{p=1}^P \sum_{i=1}^I (D_{pi} - O_{pi})^2} \quad (4.1)$$

where  $D_{pi}$  and  $O_{pi}$  are the desired and the observed values, P is number of training patterns and I is the number of output nodes.

The optimized configuration yielded a RMS error which reduced only marginally when the number of nodes in the hidden layer was increased further, but at the cost of a much longer computation time. It is worth mentioning here that while the number of nodes in the hidden layer used varied from 2 for XOR to 15 for IRIS data, the number of nodes in the hidden layer was kept same for a particular problem in different ANN algorithms to avoid any biasing towards a particular algorithm. The training of the above networks was ‘early-stopped’ to avoid any overfitting effects and this was done as soon as the RMS error reached a plateau. A maximum of  $\sim 2000$  iterations were found to be optimum for all the problems which are considered in this work. However, it is important to note that the number of iterations needed for the ANN to learn the input/output mapping depends on the complexity of the problem. In real world problems, e.g star/galaxy classification [4.6], spectra classification, primary energy estimation of Cherenkov telescopes [4.7] etc.  $\sim 10,000$  iterations have been used.

### 4.2.1 IRIS problem

The Iris flower data set or Fisher’s Iris data set is a multivariate data set introduced by [4.8] as an example of discriminant analysis. It is sometimes also called Anderson’s Iris data set because E. Anderson collected the data to quantify the morphologic variation of Iris flowers of three related species (Fig 4.1).

Fisher introduced a benchmark dataset that contains the sepal and petal measurements of different types of iris flowers. There are 150 training samples available, each of which consists of four inputs and one output. The inputs are the measured lengths and widths of the petals and the sepals, and the output is the type of iris flower, such as Setosa, Versicolor, and Virginica. The distributions of the samples with respect to the dimensions of the sepals and petals are shown in Fig 4.2 for easy visualization. It is quite evident from Fig 4.2 that the classes of Versicolor and Virginica overlap, whereas the class of Setosa is clearly separated from the other two classes.

In order to convert the training data in the  $[0,1]$  interval, all measurement values were first divided by 10. The three iris species setosa, versicolor, and virginica were



Figure 4.1: The 3 types of IRIS flowers which were classified by ANN.

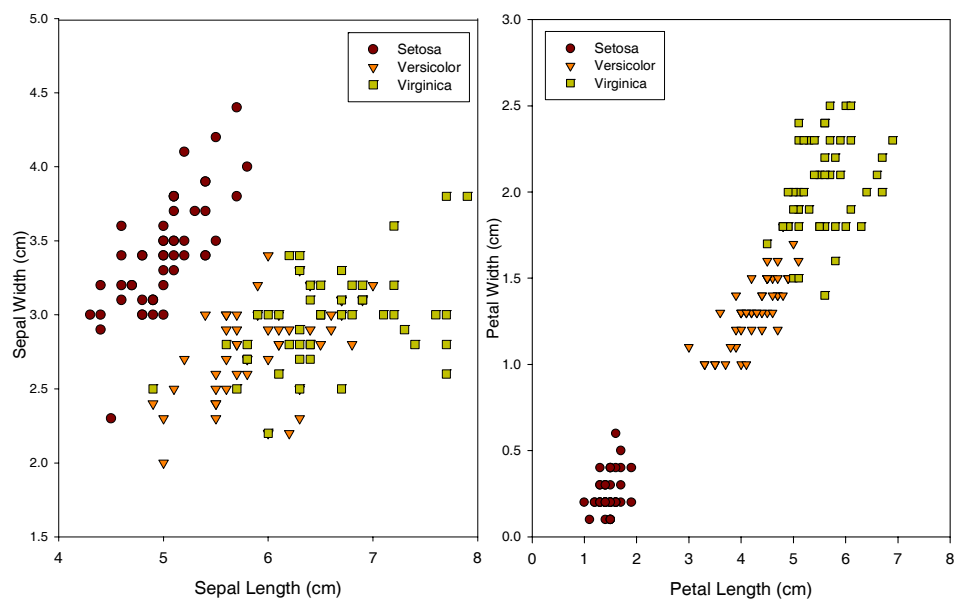


Figure 4.2: IRIS data distribution with respect to the sepal and petal lengths.

categorized with the numbers 1, 2 and 3, respectively. The configuration employed for training the ANN consists of 4:15:1 i.e, 4 neurons in the input layer, 15 neurons in the hidden layer and 1 neuron in the output layer corresponding to the category of the species. The choice of 15 neurons was found to be optimum for this task. Since the final RMS error also depends upon the choice of initial parameters (like  $\alpha$ ,  $\beta$  and the initial weights), these parameters were changed randomly 5 times and final RMS error presented here is the mean of these five RMS error values. The RMS error obtained at the end of the training process for all the algorithms alongwith time taken for completing the training is presented in Table 4.1.

<b>ANN Algorithm</b>	<b>RMS Error</b>	<b>Time (s)</b>
<i>Standard Backpropagation</i>	$2.00 \times 10^{-3}$	33
<i>Resilient Backpropagation</i>	$1.75 \times 10^{-3}$	28
<i>Lavenberg Marquardt</i>	$1.92 \times 10^{-5}$	70
<i>Congugate Gradient</i>	$9.64 \times 10^{-3}$	51
<i>Radial Basis Functions</i>	$3.99 \times 10^{-3}$	118
<i>Simulated Annealing</i>	$1.20 \times 10^{-3}$	123
<i>Neuro Fuzzy Systems</i>	$1.14 \times 10^{-3}$	311
<i>Higher order</i>	$1.21 \times 10^{-5}$	128

Table 4.1: Mean RMS error and training time of various ANN algorithms with a configuration of 4:15:1 for the IRIS problem.

The test set for the IRIS data is similar to the training set except that this data has not been presented during the training of the nets. It consists of 45 data points (15 from each class). Instead of testing the performance of all the algorithms with test data, we have only chosen one (or sometimes two) ANN algorithm for testing purpose and these are the ones which yield the lowest RMS error during their training stage. Since for the IRIS data case both Higher Order and Lavenberg-Marquardt algorithms yield the lowest (and reasonably comparable also) RMS error, we have used only these algorithms for checking their performance on the test data. The test results obtained for these algorithms suggest that while 100% classification is achievable for class Setosa, the classification for Versicolor and Virginica is only 80%. The reason for not being able to obtain 100% classification between versicolor and virginica seems to be the overlapping between these two species (Fig 4.2).

#### 4.2.2 XOR and N-bit parity problems

Exclusive disjunction, Exclusive OR or more generally XOR, is a logical operation that outputs true whenever both inputs differ (one is true, the other is false). It gains the name exclusive or because the meaning of 'or' is ambiguous when both

operands are true; exclusive or excludes that case. This is sometimes thought of as "one or the other but not both". XOR is a standard and thoroughly investigated problem in the field of neural network research. Its popularity originates from the fact that, being able to solve it was a breakthrough achieved by back-propagation algorithm, compared to the situation faced when no learning algorithm was known to solve a non linearly separable classification task such as XOR [4.9]. Apart from the XOR problem we also applied the other ANN algorithms to the generalized XOR problem i.e the N-bit parity, where the task requires to classify the sequence consisting of 1's and 0's according to whether number of 1's is odd or even [4.9]. The target for the net here is 1 or 0 depending on whether the sequence is odd or even. In the XOR problem the algorithm used has the form 2:2:1 i.e 2 neurons in the input layer, 2 neurons in the hidden layer and 1 neuron in the output layer. Also, for training the networks, more data points were also generated by incorporating random noise of 10% at the XOR inputs. The RMS error obtained for the XOR problem at the end of the training process, for all the algorithms, alongwith time taken for completing the training is presented in Table 4.2.

<b>ANN Algorithm</b>	<b>RMS Error</b>	<b>Time (s)</b>
<i>Standard Backpropagation</i>	$1.23 \times 10^{-3}$	16
<i>Resilient Backpropagation</i>	$7.72 \times 10^{-3}$	14
<i>Lavenberg Marquardt</i>	$1.59 \times 10^{-8}$	15
<i>Congugate Gradient</i>	$6.66 \times 10^{-5}$	16
<i>Radial Basis Functions</i>	$2.70 \times 10^{-3}$	18
<i>Simulated Annealing</i>	$1.18 \times 10^{-4}$	12
<i>Neuro Fuzzy Systems</i>	$2.88 \times 10^{-9}$	15
<i>Higher order</i>	$3.67 \times 10^{-7}$	28

Table 4.2: Mean RMS error and training time of various ANN algorithms with a configuration of 2:2:1 for the XOR problem.

As seen from Table 4.2, the lowest RMS error for the XOR problem is yielded by the Neuro-Fuzzy and the Lavenberg-Marquardt algorithms and hence performance testing on test data sample is done only for these two algorithms. Both these networks show  $\sim 100\%$  success rate in reproducing the XOR truth table.

The parity problem too is a demanding classification task for neural networks to solve, because the target-output changes whenever a single bit in the input vector changes. The N- bit parity consists of  $2^N$  (here  $N = 4$ ) training pairs. A 4:2:1 architecture was used by us for studying this problem. The RMS error obtained for the N-Bit problem, at the end of the training process, for all the algorithms, alongwith time taken for completing the training is presented in Table 4.3. The test set for N-Bit parity problem consists of 10 randomly generated noisy events (noise

10%). Testing of the net was done only with the Lavenberg-Marquardt algorithm network since compared to other algorithms considered in this work, it gives the lowest RMS error. The results obtained on the test data suggest that the 4-Bit parity is reproduced with an accuracy of  $\sim 90\%$ .

<b>ANN Algorithm</b>	<b>RMS Error</b>	<b>Time (s)</b>
<i>Standard Backpropagation</i>	$9.81 \times 10^{-7}$	26
<i>Resilient Backpropagation</i>	$7.12 \times 10^{-7}$	40
<i>Lavenberg Marquardt</i>	$3.43 \times 10^{-8}$	50
<i>Congugate Gradient</i>	$4.03 \times 10^{-7}$	16
<i>Radial Basis Functions</i>	$1.27 \times 10^{-4}$	28
<i>Simulated Annealing</i>	$9.01 \times 10^{-7}$	28
<i>Neuro Fuzzy Systems</i>	$3.42 \times 10^{-3}$	44
<i>Higher order</i>	$5.16 \times 10^{-4}$	55

Table 4.3: Mean RMS error and training time of various ANN algorithms with a configuration of 4:2:1 for the 4-Bit parity problem.

### 4.2.3 Two-spiral problem

Spiral curves have been the object of mathematical investigation over several hundreds of years. Many different types of spirals exist and can be described formally in different ways. The original two intertwined spirals benchmark problem was designed by Lang and Witbrock [4.10] to test the performance of classification on binary data. This particular task is difficult for most current algorithms since it requires the ANN model to learn the highly non-linear separation of the input space. In this benchmarking problem, two spirals, each of which has three complete turns, are created inside a unit square (Fig 4.3). The two-intertwined spirals problem has also been used quite extensively by other researchers as standard benchmark problem and requires the neural network to learn a mapping that distinguishes between points on two intertwined spirals. The data used by us for the 2-Spiral problem contains 194 data points (97 samples per spiral). The network configuration chosen to represent this problem has the 2:15:1 architecture, where the two inputs correspond to  $a_1$ ,  $a_2$  values of the two spirals and the 1 output corresponds to whether the value belongs to the spiral 1 or spiral 2 (1 if the point falls on one spiral and 0 if it falls on other spiral). Out of these 194 data points 164 were used for training and 30 points (15 from each spiral) were used for testing. The training results obtained for all the ANN algorithms used in the present study are presented in Table 4.4 and it is quite evident from this table that the Levenberg-Marquardt gives the best convergence results. However, from training time considerations, the standard

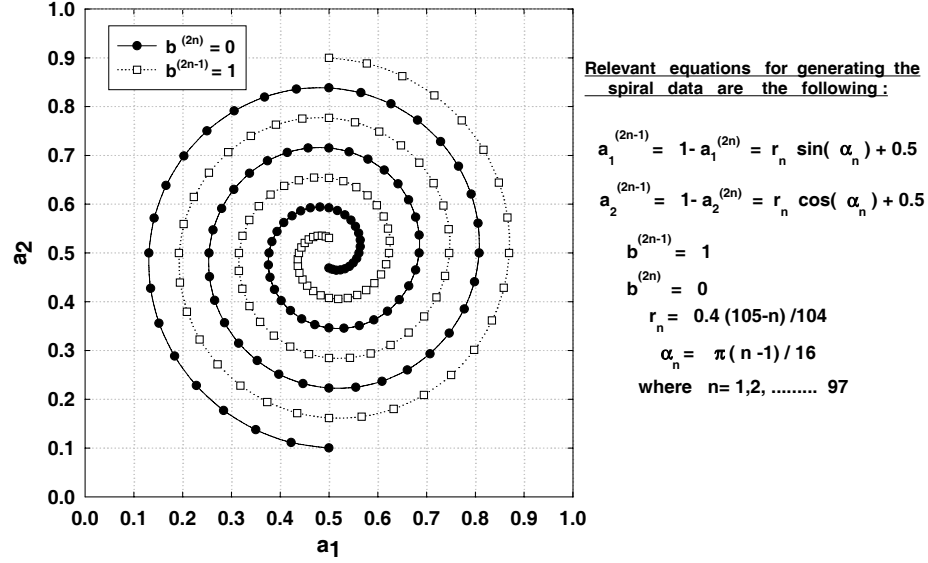


Figure 4.3: Distribution of data points on the two intertwined spirals.

backpropagation and the resilient backpropagation algorithms are seen to consume minimum training time. Performance check of the Levenberg-Marquardt algorithm, on test data for 30 points (15 from each spiral), indicates that  $\sim 70\%$  of the events are classified in the proper spiral category. This is much better as compared to resilient backpropagation which can classify only  $\sim 50\%$  of the events properly.

A consolidated report of the mean RMS error yielded by various ANN algorithms used in the present work for all the 4 benchmark problems is presented in Table 4.5. A plot of the RMS error as a function of number of iterations, for all the 4 benchmark problems is also shown in Fig 4.4 so that the performance of the backpropagation algorithm can be compared with other algorithms. For the sake of clarity, the RMS error is shown only for the Backpropagation algorithm and one more specific algorithm which yields the minimum RMS error (i.e Higher Order for IRIS, Neuro-fuzzy for XOR and Levenberg-Marquardt method for N bit parity and 2 spiral problem). It is evident from Fig 4.4 and Table 4.5 that the Lavenberg-Marquardt method yields a lower RMS error as compared to the standard backpropagation method. Even for the IRIS and XOR problems, where Higher Order and Neuro Fuzzy algorithms, respectively are found to yield the lowest RMS error, the performance of the Levenberg-Marquardt is always better than the remaining ANN algorithms. The reason for the superior performance of Levenberg-Marquardt is due to the fact that it is a combination of gradient decent and Gauss-Newton



<b>ANN Algorithm</b>	<b>RMS Error</b>	<b>Time (s)</b>
<i>Standard Backpropagation</i>	$1.51 \times 10^{-1}$	60
<i>Resilient Backpropagation</i>	$7.13 \times 10^{-1}$	56
<i>Lavenberg Marquardt</i>	$1.09 \times 10^{-1}$	110
<i>Congugate Gradient</i>	$1.23 \times 10^{-1}$	160
<i>Radial Basis Functions</i>	$1.37 \times 10^{-1}$	315
<i>Simulated Annealing</i>	$1.70 \times 10^{-1}$	200
<i>Neuro Fuzzy Systems</i>	$1.38 \times 10^{-1}$	390
<i>Higher order</i>	$1.13 \times 10^{-1}$	190

Table 4.4: Mean RMS error and training time of various ANN algorithms with a configuration of 2:15:1 for the 2-Spiral problem.

<b>ANN Algorithms</b>	<b>IRIS problem</b>	<b>XOR problem</b>	<b>N-bit parity</b>	<b>2-spiral problem</b>
<i>Backprop</i>	$2.00 \times 10^{-3}$	$1.23 \times 10^{-3}$	$9.81 \times 10^{-7}$	$1.51 \times 10^{-1}$
<i>Resilient</i>	$1.75 \times 10^{-3}$	$7.72 \times 10^{-3}$	$7.12 \times 10^{-7}$	$7.13 \times 10^{-1}$
<i>Lavenberg</i>	$1.92 \times 10^{-5}$	$1.59 \times 10^{-8}$	$3.43 \times 10^{-8}$	$1.09 \times 10^{-1}$
<i>Conjugate</i>	$9.64 \times 10^{-3}$	$6.66 \times 10^{-5}$	$4.03 \times 10^{-7}$	$1.23 \times 10^{-1}$
<i>Rad.Basis</i>	$3.99 \times 10^{-3}$	$2.70 \times 10^{-3}$	$1.27 \times 10^{-4}$	$1.37 \times 10^{-1}$
<i>Sim.Annealing</i>	$1.20 \times 10^{-3}$	$1.18 \times 10^{-4}$	$9.01 \times 10^{-7}$	$1.70 \times 10^{-1}$
<i>N.Fuzzy</i>	$1.14 \times 10^{-3}$	$2.88 \times 10^{-9}$	$3.42 \times 10^{-3}$	$1.38 \times 10^{-1}$
<i>Hig.Order</i>	$1.21 \times 10^{-5}$	$3.67 \times 10^{-7}$	$5.16 \times 10^{-4}$	$1.13 \times 10^{-1}$

Table 4.5: Comparison of mean RMS error for the different ANN algorithms considered for the study of benchmark problems.

method which combines the advantages of the local convergence properties of the Gauss-Newton method and the global properties of gradient descent. However, the computational complexity of backpropagation is only  $O(n)$  as against  $O(n^3)$  for the Levenberg-Marquardt algorithm (where  $n$  is the total number of weights in the network).

### 4.3 Application of ANN algorithms to regression problems

Artificial neural networks have become a popular tool for approximating non-linear functions in higher dimensions. Although they are not the panacea for these type

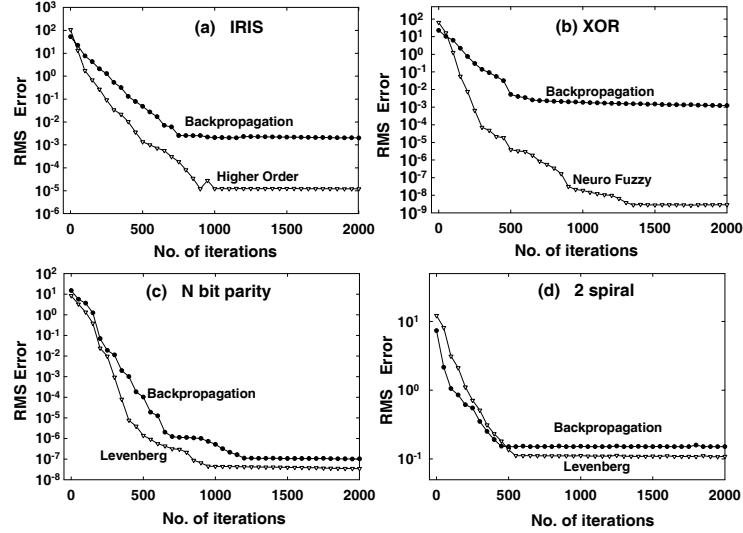


Figure 4.4: Variation of RMS error as a function of number of iterations for the 4 benchmark problems: (a) IRIS (b) XOR (c) N Bit Parity (d) 2 Spiral. The RMS error is shown only for the Backpropagation algorithm and one more specific algorithm which yields the minimum RMS error for a particular problem.

of problems, they are nevertheless recognized as a useful tool for approximating non-linear functions. Other well known methods which are conventionally used for these problems include splines [4.11], additive models [4.12], MARS [4.13], hinging hyperplanes [4.14] and CART [4.15]. While none of these methods are likely to perform consistently better than the others across a wide range of problems, it is indeed a non-trivial task to develop a method that is truly effective for all types of non linear functions. Keeping in view the superior ability of ANNs to capture non-linear behaviour of a function and its reasonably fast computation speed, we were tempted to apply the ANN as a regression tool for approximating functions like  $\cos(x)$  and a few special functions like the Gamma function, the complimentary Error function and the upper tail cumulative  $\chi^2$ -distribution function [4.16].

### 4.3.1 Approximation to $\cos(x)$

To test the performance of the ANN algorithms as a regression analysis tool, we have first applied the ANN algorithms to a simple trigonometric function like  $y=\cos(x)$ . In order to keep the output range of the network between 0 and 1, we follow the approach given in [4.9] where the function is changed to  $y = (\cos(2x)+1)/3$ . The ANN configuration chosen for this problem (i.e, 1:2:1) and the number of data points used for training ( $=200$ ) is again similar to that used by [4.9]. The training data set for this problem is synthesized by evaluating the function  $y = (\cos(2x)+1)/3$

at 200 randomly chosen points which are picked uniformly in the interval  $[0, \pi]$  range (Fig 4.5a). Additional 100 data points, following the same prescription, were also generated for testing the best ANN algorithm which produces the lowest RMS error during training. The training results obtained for all the ANN algorithms used in the present study are presented in Table 4.6. The results of the Levenberg-

<b>ANN Algorithm</b>	<b>RMS Error</b>	<b>Time (s)</b>
<i>Standard Backpropagation</i>	$9.83 \times 10^{-6}$	7
<i>Resilient Backpropagation</i>	$7.61 \times 10^{-5}$	7
<i>Lavenberg Marquardt</i>	$3.29 \times 10^{-7}$	12
<i>Congugate Gradient</i>	$7.30 \times 10^{-6}$	8
<i>Radial Basis Functions</i>	$4.15 \times 10^{-5}$	12
<i>Simulated Annealing</i>	$4.61 \times 10^{-6}$	15
<i>Neuro Fuzzy Systems</i>	$6.71 \times 10^{-6}$	25
<i>Higher order</i>	$8.12 \times 10^{-7}$	20

Table 4.6: Mean RMS error and training time of various ANN algorithms with a configuration of 1:2:1 for the  $(\cos(2x)+1)/3$  problem.

Marquardt algorithm, which yields the lowest RMS error during training, is shown in Fig 4.5b. In this figure, we have plotted the relative error in  $y$  (defined as  $(y_{ANN} - y_{EXP})/y_{EXP}$ ) as a function of  $x$  for 100 random data points generated uniformly in the interval  $[0, \pi]$ . Here, for a given value of  $x$ ,  $y_{ANN}$  is the value predicted by the ANN and  $y_{EXP}$  is expected value of the function  $(\cos(2x)+1)/3$ . It is evident from Fig 4.5b that except for  $x$  values in the vicinity of  $\pi/2$  radians, where the function  $y = (\cos(2x)+1)/3$  itself becomes close to zero, the relative error in  $y$  is in general  $<1\%$  for all other values of  $x$ .

### 4.3.2 Approximation to a few special functions

In this section, we apply the ANN algorithms as a function approximation tool to few special functions like the Gamma function, the complimentary Error function and the upper tail cumulative  $\chi^2$ -distribution function. The training and test data sets for the above special functions have been generated by using the MATHEMATICA software package.

Many important functions in applied sciences are defined via improper integrals. One of the most commonly used function amongst them is the Gamma-function. The gamma function was first introduced by the Swiss mathematician Leonhard Euler, in his goal to generalize the factorials to non integer values. The Gamma-

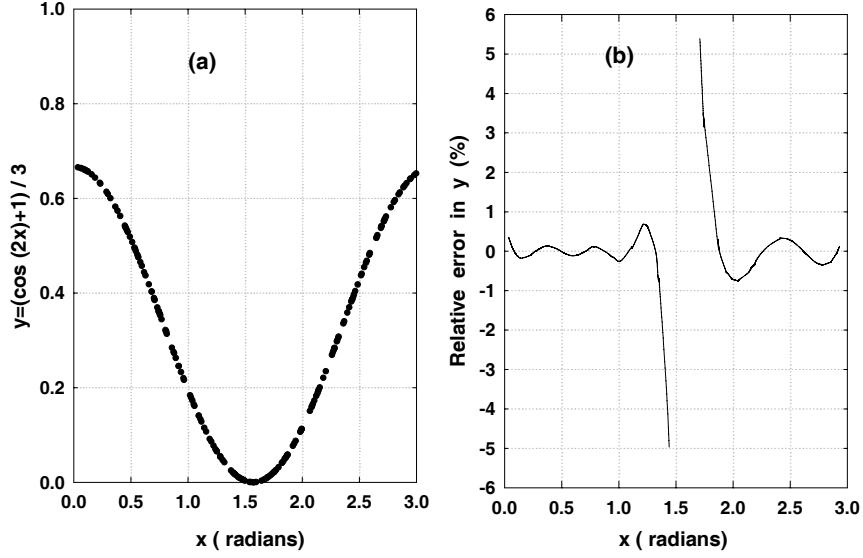


Figure 4.5: (a) The training data set for the function  $y = (\cos(2x)+1)/3$ . (b) Performance of the Levenberg-Marquardt algorithm in terms of relative error in approximating the function,  $y = (\cos(2x)+1)/3$ .

function ( $\Gamma(z)$ ) has one argument and is defined by the following integral:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \quad (4.2)$$

The approximation of the Gamma function was implemented with an ANN configuration of 1:20:1, where the one node in the input corresponds to the  $z$  value in the range  $0 < z < 20$  and the output node corresponds to  $\ln \Gamma(z)$ . The purpose of using  $\ln \Gamma(z)$  instead of  $\Gamma(z)$  directly was to avoid overflow problems even at a quite modest value of  $z$ . The training of the ANN algorithms has been done with  $\sim 1000$  values and only those values of  $z$  and  $\ln \Gamma(z)$  are used for which  $\Gamma(z) < 1.2 \times 10^{17}$ .

The second special function chosen by us to test the function approximation capability of the ANN is the complimentary function,  $\text{erfc}(x)$ . The complementary error function is an important special function which appears in the solutions of diffusion problems in heat, mass and momentum transfer, probability theory, the theory of errors and various other branches of mathematics and physics. The complimentary error function has one argument and is defined by the following integral:

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \quad (4.3)$$

Since there is direct relationship between the complimentary error function and the cumulative distribution for the Gaussian distribution, we have tried to apply the

ANN algorithms for approximating the normal tail integral. The upper tail integral, or the cumulative upper distribution function,  $Q(x)$  for Gaussian probability distribution with argument  $x$  is defined by :

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt = \frac{1}{2} \operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right) \quad (4.4)$$

The function approximation for the normal tail probability integral was implemented with an ANN configuration of 1:20:1, where the one input node corresponds to the  $x$  value ranging between 0 to 20 and the output node corresponds to  $\ln Q(x)$ . The values of  $Q(x)$  are in the range  $\sim 2.767 \times 10^{-89}$  to 0.5. About 1000 values of  $x$  and  $(\ln Q(x))$  were used for training the ANN algorithms.

The third special function chosen for testing the function approximation capability of the ANN is the cumulative distribution function of the  $\chi^2$ -probability distribution. The chi-square distribution is one of the most widely used theoretical probability distributions in inferential statistics, i.e. in statistical significance tests. The best known situations in which the  $\chi^2$ -distribution is used are the common  $\chi^2$ -tests for goodness of fit of an observed distribution to a theoretical one. The  $\chi^2$ -upper tail cumulative distribution function ( $Q(\chi^2|\nu)$ ) is defined by the following integral:

$$Q(\chi^2|\nu) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} \int_{\chi^2}^\infty e^{-t/2} t^{\nu/2-1} dt ; \quad \text{for } \nu > 0, \chi^2 \geq 0 \quad (4.5)$$

Where  $\nu$  is the degrees of freedom. The approximation of the  $\chi^2$  upper tail cumulative distribution function ( $Q(\chi^2|\nu)$ ) was implemented with an ANN configuration of 2:20:1 where the two input nodes correspond to the  $\chi^2$  and  $\nu$  values. The output node of the ANN represents the  $(\ln Q(\chi^2|\nu))$  value. About 1000 values with  $1 \leq \chi^2 \leq 100$  and  $1 \leq \nu \leq 100$  were used for training the ANN algorithms. The training of the ANN was performed with only those values of  $\chi^2$  and  $\nu$  which yield a  $Q(\chi^2|\nu)$  between  $\sim 1.757 \times 10^{-23}$  to 0.999. The results of the training regarding the mean error for all the three special functions discussed above are presented in Table 4.7.

Performance of the Levenberg-Marquardt algorithm in terms of relative error in approximating the three special functions is shown in Fig 4.6. A data sample of 100 values each is used for testing the approximation for the Gamma function (Fig 4.6a) and Gaussian upper tail probability integral (Fig 4.6b). The corresponding data samples used for testing the  $\chi^2$  upper tail probability integral is 400 for 4 different values of  $\nu$ . Referring first to the approximation of the Gamma function, it is evident from Fig 4.6a that the relative error in  $\Gamma(z)$  is  $< 0.25\%$  for  $2 < z < 20$ . However, for  $0 < z < 2$ , the relative error increases significantly to  $> \pm 0.5\%$ . Regarding the approximation of the Gaussian upper tail probability integral, the relative error in  $Q(x)$  is within  $\pm 0.3\%$  for all values of  $x$  in the range 0 to 20. The results for the approximation of  $\chi^2$  upper tail probability integral (Fig 4.6c and Fig 4.6d) indicate that the relative error in  $Q(\chi^2|\nu)$  is in general significantly more than that of the

ANN Algorithm Studied	RMS error (Gamma Function)	RMS error (Upper tail Normal dist.)	RMS error (Upper tail $\chi^2$ dist.)
<i>Backpropagation</i>	$2.25 \times 10^{-2}$	$9.16 \times 10^{-1}$	$2.91 \times 10^0$
<i>ResilientBackprop</i>	$8.97 \times 10^{-2}$	$1.51 \times 10^{-1}$	$2.01 \times 10^{-3}$
<i>Lavenberg – Marquardt</i>	$1.25 \times 10^{-6}$	$2.08 \times 10^{-9}$	$1.72 \times 10^{-5}$
<i>ConjugateGradient</i>	$5.01 \times 10^{-3}$	$2.14 \times 10^{-2}$	$4.87 \times 10^{-2}$
<i>RadialBasis</i>	$5.68 \times 10^{-3}$	$7.33 \times 10^{-3}$	$5.71 \times 10^{-2}$
<i>SimulatedAnnealing</i>	$5.23 \times 10^{-3}$	$1.52 \times 10^{-2}$	$4.20 \times 10^{-3}$
<i>Neuro – Fuzzy</i>	$4.12 \times 10^{-2}$	$8.96 \times 10^{-2}$	$1.16 \times 10^{-3}$
<i>Higher – Order</i>	$7.16 \times 10^{-5}$	$9.86 \times 10^{-8}$	$6.42 \times 10^{-5}$

Table 4.7: Mean RMS Error of various ANN algorithms with a configuration of 1:20:1 for function approximation of 3 special functions.

other 2 special functions possibly because of the presence of 2 input parameters (  $\chi^2$  and  $\nu$ ) instead of 1 as in the case of other two special functions. Keeping in view the fact that these special functions are being approximated over a very wide dynamic range ( $\sim 0.88$  to  $1.2 \times 10^{17}$  for Gamma function,  $\sim 2.7 \times 10^{-89}$  to  $0.5$  for Gaussian upper tail integral and  $\sim 1.8 \times 10^{-23}$  to  $0.999$  for  $\chi^2$  upper tail integral), we believe that the results obtained are rather encouraging. However, there is a strong need to further improve these results if one demands that approximation using ANN algorithms should yield a performance which is comparable to that of the conventional methods using numerical algorithms or other adhoc approximations. Furthermore, keeping in view the widespread use of these functions and also their highly non-linear behaviour with a very wide dynamic range, we feel that these functions can be considered as standard benchmark problems for function approximation studies using ANNs.

## 4.4 Summary

ANN algorithms have been applied to a variety of problems in various diverse areas of physics, biology, medicine, computer science etc. The main aim of most of these studies has been to use ANN-based algorithms (generally standard backpropagation) as an alternative method to conventional analysis for achieving better results. While comparative performance of some ANN algorithms like standard backpropagation, fuzzy logic, genetic algorithms, fractals etc., has been studied for various applications, a rigorous intercomparison of some of the powerful algorithms (e.g. the ones studied in this work) is still missing from the literature. The primary aim of this work has been to provide a rigorous comparative study of various powerful

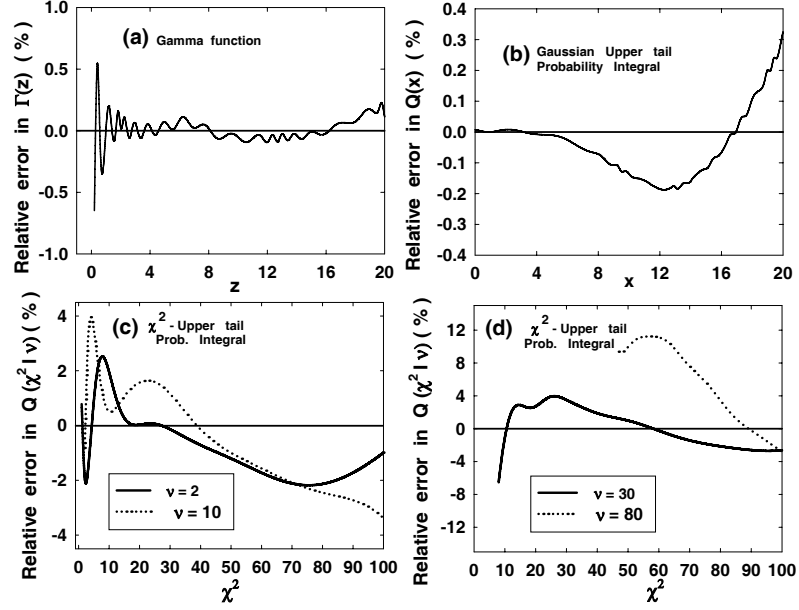


Figure 4.6: Performance of the Levenberg-Marquardt algorithm in terms of relative error in approximating the following special functions : (a) Gamma function, (b) Upper tail probability of Normal distribution, (c) and (d) Upper tail probability of  $\chi^2$  distribution for different values of  $\nu$ .

algorithms, by first applying them to standard benchmark problems and then apply them as a regression tool for approximating functions like  $\cos(x)$  and a few special functions. Our results suggest that while Levenberg-Marquardt algorithm yields the lowest RMS error for the N-bit Parity and the Two Spiral problems, Higher Order Neurons algorithm gives the best results for the IRIS data problem. The best results for the XOR problem are obtained with the Neuro Fuzzy algorithm. It is worth mentioning here that benchmark problems (IRIS, XOR/N-Bit Parity and 2-Spiral) have been also studied by numerous other workers. For example, using a 2:2:1 configuration for the XOR problem, Wang [4.9] has reported that one can achieve an accuracy of  $\sim 80\%$  with  $\sim 5000$  epochs of training. Other benchmark problems like Parity and  $\cos(x)$  have also been studied by the same author, but comparison is done only for the backpropagation and simulated annealing techniques. Likewise, the 2-Spiral problem has been studied by several workers using different algorithms like vanilla backpropagation [4.17] with configuration of 2:5:5:1, generalised regression model [4.18], vector quantization method [4.19], input coding scheme [4.20]. Complicated ANN configurations like 2:20:20:1 for the 2-Spiral problem with 50,000 training epochs and 4:4:2:1 for the IRIS problem with 30,000 training epochs has also been attempted by Lee [4.21] for achieving reasonably accurate results for these benchmark problems.

Regarding application of neural network algorithms for solving regression prob-

lems, such as evaluation of special functions like the Gamma function, the complementary Error function and the upper tail cumulative  $\chi^2$ -distribution function, we believe that such an attempt has been made in this work for the first time. The results obtained in this work indicate that, among all the ANN algorithms used in the present study, Levenberg-Marquardt algorithm yields the best results. Conventionally, two groups of approximations are found in the literature which are used for calculating these special functions. One group consists of standard numerical algorithms which, at least in theory, allow computation of the above integral to arbitrarily high precision. However, computation using these numerical algorithms requires massive computation. The second group consists of the so called 'ad hoc approximations' which require only a few carefully chosen numeric constants. Importantly, a serious limitation of most of the approximations in both the groups is that they are designed to work in a predefined range of input parameter values and the accuracy of the approximation rapidly deteriorates when the input parameters take a value outside the predefined range.

In order to appreciate the complexity of evaluating the special functions studied in this paper, it is worth discussing here some of the important approximations used for these functions. A well known method for calculating the gamma function is the so called Lanczos approximation [4.22] which computes the value of  $\Gamma(z)$  any positive real argument( $z$ ) with a high level of accuracy. Likewise, a compilation of useful approximations used for evaluating the upper tail integrals for the Gaussian and  $\chi^2$  distributions can be found in [4.23] and [4.24], respectively.

Although the comparative performance of different ANN algorithms is in general problem dependent, we feel that the study undertaken here does give an insight into the power of various ANN algorithms. Since for real world problems it is not an easy task to identify the most suitable ANN algorithm by just having a look at the problem, our results suggest that while investigating the comparative performance of other ANN algorithm, the Levenberg-Marquardt algorithm deserves a serious consideration and cannot be rejected outright because of its training time overheads.

---

---

## 4.5 References

- [4.1] M.D. Christopher et al., World Congress on Computational Intelligence IEEE 1213 (1994).
- [4.2] K. Funahashi, Neural Networks 2 (1989) 183.
- [4.3] K. Hornik et al, Neural Networks 2 (1989) 359.
- [4.4] R.K. Bock et al., Nucl. Instr and Meth. A 516 (2004) 511.



- [4.5] V.K. Dhar et al., *Pramana Journal of Physics*, 74 (2010) 307.
- [4.6] H.P. Singh et al., *MNRAS* 295 (1998) 312.
- [4.7] V.K. Dhar et al., *Nucl. Instr. and Meth.A* 606 (2009) 795.
- [4.8] R.A. Fisher et al., *Annals of Eugenics* 7 (1936) 179.
- [4.9] X. Wang et al., *Trans. of IEEJ* 124 (2004) 842.
- [4.10] K. Lang et al., *Proc. Connectionist models*, Morgan Kaufmann Publishers Inc (1998) 52.
- [4.11] D. Barry, *Nonparametric Bayesian Regression*, *Ann. Statist.* 14 (1986) 934.
- [4.12] T. Hastie et al., *Statist. Science* 1 (1986) 297.
- [4.13] J.H. Friedman, *Tech. Rep. 102*, Lab. for Computational Statistics, Dept. of Statistics, Stanford University (1990).
- [4.14] L. Breiman, *IEEE Trans. Information Theory* 39 (1993) 999.
- [4.15] L. Breiman et al., *Classification and Regression Trees*, Wadsworth and Brooks, Pacific Grove CA, (1984).
- [4.16] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, US Government printing office Washington D.C (1972).
- [4.17] G.A. Carpenter et al., *IEEE Trans. on Neural Networks* 3 (1992) 698.
- [4.18] E.W.M. Lee et al., *Advanced Engineering Informatics* 20 (2006) 213.
- [4.19] T. Denoeux et al., *Neural Network* 6 (1993) 351.
- [4.20] J.C. Jia et al., *Electronic Letters* 31 (1995) 1267.
- [4.21] J. Lee, *Proc. International Joint Conference on Neural Networks* 1 (2003) 410.
- [4.22] C. Lanczos, *SIAM Journal on Numerical Analysis*, 1 (1964) 86.
- [4.23] W. Bryc, *Appl. Math. Comput.* 127 (2002) 365.
- [4.24] L. Canal, *Computational Statistics and Data Analysis* 48 (2005) 803.

# Chapter 5

## ANN based gamma-hadron segregation for TACTIC telescope

### 5.1 Introduction

Gamma-ray photons in the TeV energy range (0.1-50TeV), are expected to come from a wide variety of cosmic objects within and outside our galaxy. While this promise of the cosmic TeV  $\gamma$ -ray probe has been appreciated for quite long, it was the landmark development of the imaging technique and the principle of stereoscopic imaging, proposed by Whipple [5.1] and the HEGRA [5.2] groups, respectively, that revolutionized the field of ground-based very high-energy (VHE)  $\gamma$ -ray astronomy.

The success of VHE  $\gamma$ -ray astronomy, however depends critically on the efficiency of  $\gamma$ /hadron classification methods employed. Thus, in order to improve the sensitivity of ground based telescopes, the main challenge is to improve the existing  $\gamma$ /hadron segregation methods to efficiently reduce the background cosmic-ray contamination and at the same time also retain higher number of  $\gamma$ -ray events. Detailed Monte-Carlo simulations, pioneered by Hillas [5.3], show that the differences between Cherenkov light emission from air showers initiated by  $\gamma$ -rays and protons (and other cosmic-ray nuclei) are quite pronounced, with the proton image being broader and longer as compared to the  $\gamma$ -ray image. This led to the development and successful usage of several image parameters in tandem, a technique referred to as the Supercuts/Dynamic Supercuts method. Although the efficiency of this  $\gamma$ /hadron event classification methodology, has been confirmed by the detection of several  $\gamma$ -ray sources by various independent groups including us, there is a need to search for still more sensitive/efficient algorithms for  $\gamma$ /hadron segregation. The conventionally used Supercuts/Dynamic Supercuts method, though using several image parameters simultaneously, with some of them also being energy dependent, is still a one dimensional technique, in the sense that the parameters it uses for classification are treated separately and the possible correlations among the parameters are ignored.

The multivariate analysis methods, proposed by various groups, for discriminating between  $\gamma$ -rays and hadrons are the following: Multidimensional Analysis based on Bayes Decision Rules [5.4], Mahalanobis Distance [5.5], Maximum Likelihood [5.6], Singular Value Decomposition [5.7], Fractals and Wavelets [5.8-5.9] and Neural Networks [5.10-5.11]. The comparative performance of different multivariate classification methods like Regression ( or Classification) trees, kernel methods, support vector machines, composite probabilities, linear discriminant analysis and Artificial Neural Networks (ANN) has also been studied by using Monte Carlo simulated data for the MAGIC telescope. A detailed compilation of this study is reported in [5.12]. The results published in the above work indicate that while as the performance of Classification Trees, Kernel and Nearest-Neighbour methods are very close to each other, the different ANN methods employed (feed-forward, random search and multilayer perceptron) yield results over a wide range. The feed-forward method gives a significance of  $\sim 8.75 \sigma$ , whileas multilayer perceptron gives a somewhat poorer significance of  $\sim 7.22\sigma$  [5.12]. The discrimination methods like Linear Discriminant Analysis and Support Vector Machines are found to be inferior compared to others [5.12].

## 5.2 Brief description of some similar applications using ANN

Research activity in the last decade or so has established that ANN based algorithms are promising alternatives to many conventional classification methods. The advantages of ANN over the conventionally used methods are mainly the following: Firstly, ANN are data driven, self- adaptive methods, since they adjust themselves to given data without any explicit specification of the functional form for the underlying model. Secondly, they are universal function approximators as they can approximate any function with an arbitrary accuracy [5.13]. Third and most important, ANN are able to estimate the posterior probability which provides the basis of establishing classification rule and performing statistical analysis [5.14,5.15]. These statistical methods, though important for classification are merely based on bayesian decision theory in which posterior probability plays a central role. The fact that ANN can provide an estimate of posterior probability implicitly establishes the strong connection between the ANN and statistical methods. A direct comparison between the two, however, is not possible as ANN are non-linear and model free methods, while as statistical methods are mostly linear and model based.

Several  $\gamma$ -ray astronomy groups have already explored the feasibility of using ANN for  $\gamma$ / hadron separation work. While nobody has so far worked with primary ANN (i.e using Cherenkov images directly as inputs to ANN), the results reported are mainly from the use of secondary ANN where various image parameters are used as inputs to the ANN. In an attempt to examine the potential of ANN for

improving the efficiency of the imaging technique,  $\gamma$ -ray and proton acceptance of  $\sim 40\%$  and  $\sim 0.7\%$ , respectively was achieved by Vaze [5.16] by using 8 image parameters as inputs to the ANN. A detailed study of applying ANN to imaging telescope data was attempted by Reynolds and Fegan [5.11] and results of their study indicate that the ANN method although being superior to other methods like maximum likelihood and singular value decomposition does not yield better results than the Supercuts Method. The work reported by Chilingarian in [5.10] by using 8 image parameters as inputs to the ANN, on the other hand, indicates a slightly better performance of the ANN method as compared to the Supercuts procedure. Using a network configuration of 4:5:1 on the Whipple 1988-89 Crab Nebula data, the author has reported only marginal enhancement in the statistical significance ( viz.,  $\sim 35.80\sigma$  as against  $\sim 34.30\sigma$  obtained with the Supercuts method), but there is a significant increase in the number  $\gamma$ -rays retained by the ANN ( viz.,  $\sim 3420$  as against  $\sim 2686$  obtained with the Supercuts method). Application of Fourier transform to Cherenkov images and then using the resulting spatial frequency components as inputs to a Kohonen unsupervised neural network for classification has been reported by Lang [5.17]. The performance of Multifractal and Wavelet parameters was examined by the HEGRA collaboration in [5.18] by using a data sample from the Mkn 501 observation. The authors of the above work report that combining Hillas and multifractal parameters using a neural network yields a slight improvement in performance as compared to the Hillas parameters used alone.

There are also many other assorted [5.19-5.20] and non-imaging applications including data collected by extensive air shower arrays where ANN have been applied. Bussino and Mari [5.21] employed a backpropagation based ANN model for separating electromagnetic and hadronic showers detected by an air shower array. They achieved a  $\sim 75\%$  identification for  $\gamma$ -rays and  $\sim 74\%$  identification for protons. Maneva et al [5.22] used a ANN algorithm for the CELESTE data. Dumora et al [5.23] have also reported promising results for CELESTE data where ANN method was used for discriminating the  $\gamma$  /hadron Cherenkov events for the wavefront sampling telescope. The standard Stuttgart Neural Network Simulator (SNNS) package has also been used for  $\gamma$ /hadron segregation for the data obtained from AGRO-YBJ experiment [5.24]. Application of backpropagation based ANN method for separating  $\gamma$ /hadron events recorded by the HEGRA air shower array has been studied by Westerhoff et al [5.25].

Keeping in view the encouraging results reported in the above cited literature, in particular the results published in [5.10-5.12], we studied the  $\gamma$ / hadron segregation potential of various ANN algorithms, by applying them to the Monte Carlo simulated data. The idea of applying ANN for determining the energy of the  $\gamma$ -rays, from a point source, has already been used by us [5.26] for determining the energy spectra of the Crab Nebula, Mrk421 and Mrk501, as measured by the TACTIC telescope, details of which will be discussed in the following chapter.

## 5.3 Statistical analysis of Cherenkov image parameters

The details of the procedure related to generation of the simulated data base for the TACTIC telescope which has been used in the present study are discussed in chapter 2 and the Cherenkov imaging parameters used here are discussed in chapter 1.

### 5.3.1 Statistical analysis of various parameters for selecting the optimal features

The success of any classification technique depends on the proper selection of the variables which are to be used for the event segregation and the agreement between the expected and the actual distributions of these variables. Fig 5.1, shows the distributions of the Cherenkov image parameters LENGTH, WIDTH, DISTANCE and  $\alpha$  for simulated protons events and for the actual Cherenkov events recorded by the telescope. The data plotted here has been first subjected to pre-filtering cuts with  $SIZE \geq 50$  photoelectrons (pe) and  $(0.4^0 \leq DISTANCE \leq 1.4^0)$  in order to ensure that the events recorded are robust and well contained in the camera. The simulated image parameter distribution of  $\gamma$ -rays has also been shown in the figure below for comparison. The observed image parameter distributions are found to closely match the distributions obtained from simulations for proton-initiated showers, thus suggesting that the response of the telescope is reasonably close to that predicted by simulations. For converting the event SIZE, recorded in charge to digital counts, to corresponding number of photoelectrons, we have used a conversion factor of  $1\text{pe} \cong 6.5$  counts [5.27]. In order to understand and improve upon the existing  $\gamma$ /hadron segregation methods it is important to estimate the discriminating capability of each of the Cherenkov image parameters and their correlations[5.4]. The image parameters considered for this correlation study are : SIZE, LENGTH, WIDTH, DISTANCE, FRAC2 and  $\alpha$ .

In order to select image parameters which are best suited for  $\gamma$ / hadron separation we have applied the following tests: Student's t-test, Welch's t-test, Mann Whitney U-test (also known as Wilcoxon rank-sum test) and the Kolmogorov - Smirnov test (KS test) [5.28]. The Student's t-test and Welch's t-test belong to the category of parametric tests which assume that the data are sampled from populations that follow a Gaussian distribution. While as, the Students unpaired t-test assumes that the two populations have the same variances, the Welch's t-test is a modification of the t- test which does not assume equal variances. Tests that do not make any assumptions about the population distribution are referred to as nonparametric tests. Mann Whitney U-test and Kolmogorov - Smirnov test belong to this category of tests. While the nonparametric tests are appealing because they make fewer assumptions about the distribution of the data, they are less powerful

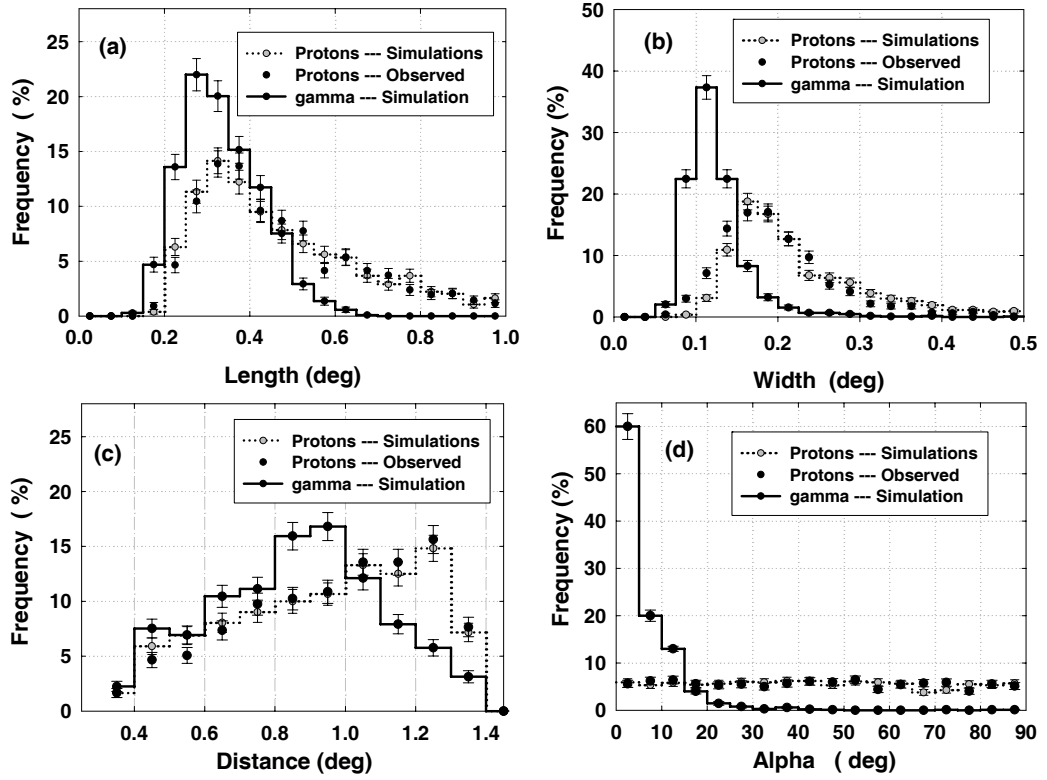


Figure 5.1: Comparison of image parameter distributions (a) LENGTH, (b) WIDTH, (c) DISTANCE and (d) Alpha ( $\alpha$ ) from real and the Monte Carlo simulated data for proton events. The simulated image parameter distribution of  $\gamma$ -rays has also been shown in the figure for comparison.

than the parametric tests. This means that the corresponding probability values tend to be higher, making it harder to detect real differences as being statistically significant. When large data samples are considered, the difference in power is minor. Furthermore, it is worth mentioning here that the parametric tests are robust to deviations from Gaussian distributions, so long as the samples are large.

In order to apply the above mentioned tests to simulated data of  $\gamma$ -ray and proton initiated showers we have used  $\sim 6000$  events each, at a zenith angle of  $25^\circ$  and the results of these one-dimensional tests are summarized in Table 5.1.

Since the P-values (i.e the probability of rejecting the null hypothesis that the  $\gamma$ -ray data sample and the proton-data sample come from the same population) are usually very small we have instead used the value of the corresponding statistic for rejecting or accepting the null hypothesis. In other words t-statistic values are

Table 5.1: Statistic values of various parametric and non-parametric statistical tests. Larger value of the statistic indicates that the corresponding probability of rejecting the null hypothesis, that the  $\gamma$ -ray data sample and the proton-data sample come from the same population, is low.

Notations used are SIZ=SIZE, LEN=LENGTH, WID=WIDTH, DIS=DISTANCE, FR2=FRAC2.

	<b>Student's t-test</b>	<b>Welch's t-test</b>	<b>Mann-Whitney U-test</b>	<b>KS D-test</b>
	<b>t</b>	<b>t</b>	<b>z</b>	<b>D</b>
<b>SIZ</b>	1.95	1.94	8.66	0.09
<b>LEN</b>	138.80	138.75	90.20	0.85
<b>WID</b>	120.96	120.28	84.75	0.76
<b>DIS</b>	19.65	19.64	17.18	0.18
<b>FR2</b>	200.84	200.94	92.69	0.90
<b><math>\alpha</math></b>	112.57	112.53	82.89	0.76

given in the Table 5.1 for expressing the results of Student's t-test and Welch's t-test. Similarly, for Mann-Whitney U test the z-statistic values are given in the table (where  $z = (U - m_U)/\sigma_U$  with  $m_U$  and  $\sigma_U$  as the mean and the standard deviation of U). For the Kolmogorov-Smirnov test we have calculated D-statistic (i.e. maximum vertical distance between the two cumulative frequency distributions). On examining Table 5.1 it is evident that four image parameters (viz., LENGTH, WIDTH, FRAC2 and  $\alpha$ ) have a significant potential of providing efficient  $\gamma$ /hadron separation. Larger the value of the corresponding statistic, lower is the corresponding probability of rejecting the null hypothesis that the  $\gamma$ -ray data sample and the proton-data sample come from the same population.

In order to estimate the statistical relationship between two image parameters for  $\gamma$ -ray data sample and the proton-data samples separately we have also calculated the Pearson product-moment correlation coefficient. Following the standard procedure, it is obtained by dividing the covariance of the two variables by the product of their standard deviations. The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables. The results of this study, obtained separately for  $\gamma$ -ray and proton-data samples, are presented in Tables 5.2 and Table 5.3, respectively.

The values of the t-statistic corresponding to each correlation coefficient are also given in these Tables (numbers within parentheses). These values can be used for assessing the significance of the correlation. Larger value of the z-statistic indicates that the corresponding probability of rejecting the null hypothesis that the observed value comes from a population in which correlation coefficient  $\sim 0$ , is low. If the

Table 5.2: Correlation matrix for simulated  $\gamma$ -ray data sample at a zenith angle of  $25^\circ$ . The values listed below for each correlation coefficient (numbers within parentheses) are the corresponding z-statistic values obtained using Fisher transformation.

	<b>SIZ</b>	<b>LEN</b>	<b>WID</b>	<b>DIS</b>	<b>FR2</b>	<b><math>\alpha</math></b>
<b>SIZ</b>	1.000 (- - -)	0.394 (33.206)	0.474 (41.692)	0.072 (5.603)	-0.441 (38.051)	-0.037 (2.881)
<b>LEN</b>	0.394 (33.206)	1.000 (- - -)	0.615 (60.452)	0.038 (2.908)	-0.709 (78.069)	0.196 (15.466)
<b>WID</b>	0.474 (41.692)	0.615 (60.452)	1.000 (- - -)	-0.396 (33.360)	-0.569 (53.680)	0.456 (39.649)
<b>DIS</b>	0.072 (5.603)	0.038 (2.908)	-0.396 (-33.360)	1.000 (- - -)	-0.034 (-2.615)	-0.366 (30.491)
<b>FR2</b>	-0.441 (38.051)	-0.709 (78.069)	-0.569 (53.680)	-0.034 (-2.61)	1.000 (- - -)	-0.064 (4.927)
<b><math>\alpha</math></b>	-0.037 (2.881)	0.196 (15.466)	0.456 (39.649)	-0.366 (30.491)	-0.064 (4.927)	1.000 (- - -)

correlation coefficient is  $\rho$  the Fisher transformation can be defined as:

$$z = \frac{1}{2} \ln \left( \frac{1 + \rho}{1 - \rho} \right) \quad (5.1)$$

The Fisher  $\rho$ -to- $z$  transformation [5.29] has also been applied to assess the significance of the difference between two correlation coefficients (say  $\rho_1$  and  $\rho_2$ ) found in two independent samples. The relevant expression to calculate this is given by :

$$z_{12} = \left( \frac{|\rho_1 - \rho_2|}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \right) \quad (5.2)$$

where  $\rho_1$  and  $\rho_2$  are the two correlation coefficients,  $n_1$  and  $n_2$  are respectively the number of data points used while calculating  $\rho_1$  and  $\rho_2$ . Table 5.4 gives the values for the Fisher matrix of various image parameters for the simulated  $\gamma$ /proton sample.

On examining Tables 5.2, 5.3 and 5.4, one can select the image parameters for achieving optimum  $\gamma$ /hadron segregation. This can be done on the basis of identifying parameters for which the difference between their correlation coefficients is maximum. As seen in Table 5.4, WIDTH- $\alpha$  pair yields the largest Fisher test value. Furthermore, it is also encouraging to find that the other well known characteristics of Cherenkov image parameters are in good agreement with our results. For



Table 5.3: Correlation matrix for simulated proton data sample at a zenith angle of  $25^\circ$ . The values listed below for each correlation coefficient (numbers within parentheses) are the corresponding z-statistic values (obtained using Fisher transformation).

	<b>SIZ</b>	<b>LEN</b>	<b>WID</b>	<b>DIS</b>	<b>FR2</b>	<b><math>\alpha</math></b>
<b>SIZ</b>	1.000 (--)	0.036 (2.757)	0.273 (21.950)	-0.301 (2.332)	-0.083 (6.472)	-0.008 (0.624)
<b>LEN</b>	0.036 (2.757)	1.000 (--)	0.360 (29.916)	-0.023 (1.792)	-0.618 (60.947)	0.086 (6.691)
<b>WID</b>	0.273 (21.950)	0.360 (29.916)	1.000 (--)	-0.036 (2.806)	-0.510 (46.998)	0.005 (0.368)
<b>DIS</b>	-0.301 (2.332)	-0.023 (1.792)	-0.036 (2.806)	1.000 (--)	-0.006 (0.481)	0.012 (0.958)
<b>FR2</b>	-0.083 (6.472)	-0.618 (60.947)	-0.510 (45.998)	-0.006 (0.481)	1.000 (--)	-0.028 (2.154)
<b><math>\alpha</math></b>	-0.008 (0.624)	0.086 (6.691)	0.005 (0.368)	0.012 (0.958)	-0.028 (2.154)	1.000 (--)

example, dependence of the image shape parameters (i.e LENGTH and WIDTH) on SIZE for  $\gamma$ -rays. Both these parameters yield positive correlation coefficient of  $\sim 0.394$  and  $\sim 0.474$  as shown in Table 5.2. Since SIZE parameter of an image provides an approximate estimate of the  $\gamma$ -ray primary energy both these parameters are expected to be correlated with the event SIZE. The modification of the Supercuts procedure to Dynamic (or extended) Supercuts follows the same principle. Negative correlation between DISTANCE and  $\alpha$  for  $\gamma$ -rays coming from a point source is also seen in Table 5.2 in accordance with the expected relationship between these image parameters. Thus, on the basis of results presented in Tables 5.2, 5.3 and 5.4, one can confidently say that there is sufficient scope for utilizing the differences in the correlation between various image parameters for developing alternate  $\gamma$ /hadron segregation methodologies.

Keeping in view the fact that, for proton initiated showers (as also in general for other cosmic-ray primaries), the image parameter  $\alpha$  is expected to be independent of other image parameters because of the isotropic nature of the cosmic-rays we will not use it in the ANN-based  $\gamma$ /hadron segregation methodology. Justification for following this approach is also evident in Table 5.3, where for the proton data sample, one finds negligible correlation between  $\alpha$  and other image parameters. Thus, for extracting the  $\gamma$ -ray signal from the cosmic-ray background, we will use the frequency distribution of the  $\alpha$  parameter for the ANN selected events. The distribution is expected to be flat for cosmic-rays and should reveal a peak at smaller  $\alpha$  values for  $\gamma$ -rays coming from a point source. In all, we will use the following six

Table 5.4: Fisher Matrix for the simulated  $\gamma$ /hadron data sample at a zenith angle of  $25^\circ$ . The matrix can be used to assess the significance of the difference between two correlation coefficients.

	<b>SIZ</b>	<b>LEN</b>	<b>WID</b>	<b>DIS</b>	<b>FR2</b>	<b><math>\alpha</math></b>
<b>SIZ</b>	--	20.83	12.87	5.6	21.30	1.59
<b>LEN</b>	20.83	--	18.65	3.32	8.95	6.16
<b>WID</b>	12.87	18.65	--	20.97	4.56	26.70
<b>DIS</b>	5.60	3.32	20.97	--	1.51	21.68
<b>FR2</b>	21.30	8.95	4.56	1.51	--	1.96
<b><math>\alpha</math></b>	1.59	6.16	26.7	21.68	1.96	--

image parameters in the ANN-based  $\gamma$ /hadron segregation methodology : Zenith angle ( $\theta$ ), SIZE, LENGTH, WIDTH, DISTANCE and FRAC2. Use of  $\theta$  angle as an additional variable can be justified by keeping in view the fact that as  $\theta$  angle increases, the line of sight distance to the shower maximum also increases, making all projected dimensions of the shower (i.e, LENGTH and WIDTH) smaller. The shape parameters LENGTH and WIDTH are expected to approximately scale as  $\propto \cos(\theta)$ .

## 5.4 ANN methodology

The ANN methodology along with the details and power of the various ANN algorithms considered by us have been discussed in detail in chapter 3 of this thesis. It is however important to mention here that for real world problems, the definitions presented in chapter 3, serve only as a guideline and the actual performance of the ANN models on real world problems does not necessarily follow these theoretical predictions. Thus, the varied algorithms under the ANN domain can not be used as off the shelf algorithms until sufficient expertise in the field is obtained. There are several other issues involved in designing and training a multilayer neural network. These are : (a) Selecting appropriate number of hidden layers in the network; (b) Selecting the number of neurons to be used in each hidden layer; (c) Finding a globally optimal solution that avoids local minima; (d) Converging to an optimal solution in a reasonable period of time; (e) Overtraining of the network and (f) Validating the neural network to test for overfitting.

It is important to mention that while lot of emphasis has been put lately on the use of Random Forest (RF) technique as an efficient tool for  $\gamma$ -hadron segregation, we believe that a properly selected and well trained neural net algorithm is equally as efficient for this purpose. The results obtained by [5.12] in their study obtained a Quality factor (QF) of  $\sim 2.8$  and  $\sim 3.0$  for Random Forest and ANN methods respectively when applied to the MAGIC data. The maximum significance also

turns out to be comparable at  $\sim 8.74\sigma$  and  $\sim 8.75\sigma$  for RF and ANN respectively. In another study conducted by Boinee et al. [5.30] on the MAGIC Cherenkov telescope experiment, detailed comparison of RF, ANN, Support Vector Machines and Classification Trees have been presented. While as, the optimized RF technique resulted in a classification accuracy of  $\sim 81.24\%$ , the classification accuracy for ANN turned out to be  $\sim 81.75\%$  with a mean error rate of  $\sim 0.276$  and  $\sim 0.256$  for the Random Forest and ANN techniques respectively, thereby suggesting that the two techniques are at best comparable. The results obtained from other methods turn out to be quite inferior compared to ANN and Random Forest, suggesting that both these methods are equally suitable.

## 5.5 Gamma/hadron separation using ANN

### 5.5.1 Preparation of training, testing and validation data

Training the ANN means iteratively minimizing the error between the desired output and the ANN generated value, with respect to the network weights. Clearly, in order for the network to yield appropriate outputs for given inputs, the weights must be set to suitable values. This is done by 'training' the network on a set of input vectors, for which the ideal outputs (targets) are already known. For training the ANN we have used  $\sim 13,750$   $\gamma$ -ray simulated events (details presented in chapter 2) following a power law distribution with a differential spectral index of  $\sim -2.6$ . This data-base was obtained by combining together  $\sim 2,750$  events each at 5 different zenith angles ( $\theta = 5^\circ, 15^\circ, 25^\circ, 35^\circ$  and  $45^\circ$ ). The cosmic-ray data of  $\sim 11,290$  events, used for training the ANN, is the actual experimental data recorded by the TACTIC telescope and was prepared in the following manner. Around one-third of the data used ( $\sim 3,163$  events) were recorded in the Crab Nebula off source direction. From the Crab Nebula on-source data base, collected between Nov.10, 2005 - Jan. 30, 2006, we used another ( $\sim 3163$  events) for which  $\alpha \geq 27^\circ$  and are hence certainly cosmic-ray events. The remaining one-third portion of the data was taken from  $\sim 30$ h of Mrk 421 off-source observations and this data was collected during the same observing season. The zenith angle of the off-source observation was restricted to  $\leq 45^\circ$ . The reason for generating the training data in this manner was to ensure that all possible systematic influences on the training of the network such as variable sky brightness in different directions are also included during the training procedure. Using the experimental data-base for the protons is a useful way of training, since it helps ANN to recognize the latent patterns, if any, in a better way which can otherwise be difficult to replicate in simulations e.g, in situations when the sky brightness is higher than what has been assumed in simulations. The importance of using real background hadronic events instead of simulated events has also been demonstrated in [5.11].

The test data set consists of an independently generated sample of about 44,831

events (mixture of  $\sim 24,603$  simulated  $\gamma$ -ray and  $\sim 20,228$  actual cosmic-ray events), which has not been used while training the ANN. This data set has exactly the same format as the training data set and is generated in the same manner as the training data. A validation data sample of  $\sim 29,798$  events (mixture of 16,424 simulated  $\gamma$ -ray and 13,374 actual cosmic-ray events) is used for verifying that the network retains its ability to generalize and is not ‘over-trained’.

### 5.5.2 ANN training and optimizing the hidden layer nodes

The network used in this work comprises 6 nodes in the input layer with one node each for Zenith angle ( $\theta$ ), SIZE, LENGTH, WIDTH, DISTANCE and FRAC2 and one neuron in the output layer whose value decides the class to which the output is to be categorized. This value is designated as 0.1 or 0.9 depending upon whether the event in question is a  $\gamma$ -ray or a cosmic-ray event respectively. In order to determine the optimum number of neurons in the hidden layer we evaluate the Mean Square Error (MSE) generated by the network. The MSE for the network is defined as:

$$MSE = \frac{1}{2} \frac{1}{PI} \sum_{p=1}^P \sum_{i=1}^I \left( \frac{D_{pi} - O_{pi}}{D_{pi}} \right)^2 \quad (5.3)$$

where  $D_{pi}$  and  $O_{pi}$  are the desired and the observed values and P is the number of training patterns and I is the number of outputs, which happens to be 1 in our case. Thus MSE defined above, is the sum of the squared differences between the desired output and the actual output of the output neurons averaged over all the training exemplars [5.31]. The ANN algorithms used in the present work are the following: Backpropagation, Resilient Backpropagation, Backprop-momentum, Conjugate Gradient, One Step Secant, Higher Order Neurons, Levenberg-Marquardt and the Neuro Fuzzy.

With regard to choosing the number of nodes in the hidden layer, it is well known that, while using too few nodes will starve the network of the resources that it needs to solve a particular problem; choosing too many nodes has the risk of potential overfitting where the network tends to remember the training cases instead of generalizing the patterns. In order to find the optimum number of nodes in the hidden layer we employed a two step procedure. In the first step we varied the number of nodes in the hidden layer from 5 to 60 (in steps of 5 up to 40 and in steps of 10 thereafter) and noted down the MSE for each of the configurations. In the second step, we deliberately used significantly higher number of nodes in the hidden layer (equal to 90) and then applied the Singular Value Decomposition (SVD) method for identifying the redundant nodes [5.19, 5.32-5.34]. It is worth mentioning here that determining the optimum number of neurons in the hidden layer by sequentially increasing the number of nodes from 60 onwards involves massive computational effort, hence the need of applying the SVD method is justified.

In the SVD method, the weight matrix (denoted by F in the present work) was generated by finding the output of each of the 90 nodes before subjecting them to the nonlinear transformation (i.e output of the hidden node). With a total of 25,040 training patterns and one hidden layer with 90 nodes, the matrix F has thus 25,040 rows and 90 columns. The SVD of the matrix F is given by  $F=U S V^T$ , where U and V are the orthogonal matrices and S is a diagonal matrix with 25,040 rows and 90 columns. The matrix S contains the singular values of F on its diagonal. The dominance of the significant singular values of F (say g out of a total p singular values) is found out by using the so called percentage of energy explained ( $P_{ex}$ ) and is defined as :

$$P_{ex} = \frac{\sum_{i=1}^g S_i^2}{\sum_{i=1}^p S_i^2} \times 100 \quad (5.4)$$

where  $S_1, S_2, S_3, \dots, S_p$  are the singular values of F arranged in their descending order [5.35]. The results of this study are shown in Fig 5.2, where  $P_{ex}$  is plotted as a function of number of nodes in the hidden layer for a representative example of 4 ANN algorithms. Consolidated results concerning the performance of the various algorithms with regard to their corresponding MSE values for the training, test and validation data samples are given in Table 5.5. The results presented in this table shown separately for 35 and 90 nodes in the hidden layer, can be used for checking whether the ANN algorithm is "over-trained" or not. When the network is over-trained, the MSE for the test and validation data samples are expected to be significantly higher than the corresponding value of MSE achieved during training.

Table 5.5: MSE values of various ANN algorithms for the training, test and validation data samples. The two values presented in the table correspond to 35 and 90 nodes in the hidden layer.

<b>Algorithm</b>	<b>Train 35/90</b>	<b>Test 35/90</b>	<b>Valid 35/90</b>
<i>Backpropagation</i>	0.103/0.102	0.103/0.103	0.103/0.103
<i>BackpropMomentum</i>	0.156/0.158	0.157/0.159	0.156/0.158
<i>ResilientBackprop</i>	0.035/0.033	0.036/0.035	0.036/0.034
<i>ScaleConjugate</i>	0.047/0.040	0.046/0.041	0.047/0.041
<i>OneStepSecant</i>	0.053/0.050	0.053/0.051	0.053/0.051
<i>LavenbergMarquardt</i>	0.017/0.015	0.017/0.030	0.017/0.031
<i>HigherOrder</i>	0.039/0.033	0.040/0.033	0.040/0.034
<i>NeuroFuzzy</i>	0.062/0.062	0.062/0.063	0.062/0.062

The optimum number of nodes for  $P_{ex} \sim 99.9 \%$  is also marked in the figures by full vertical lines. For  $P_{ex} \sim 99.9 \%$ , one can easily find from the this figure that the optimum number of nodes needed for obtaining the desired results varies

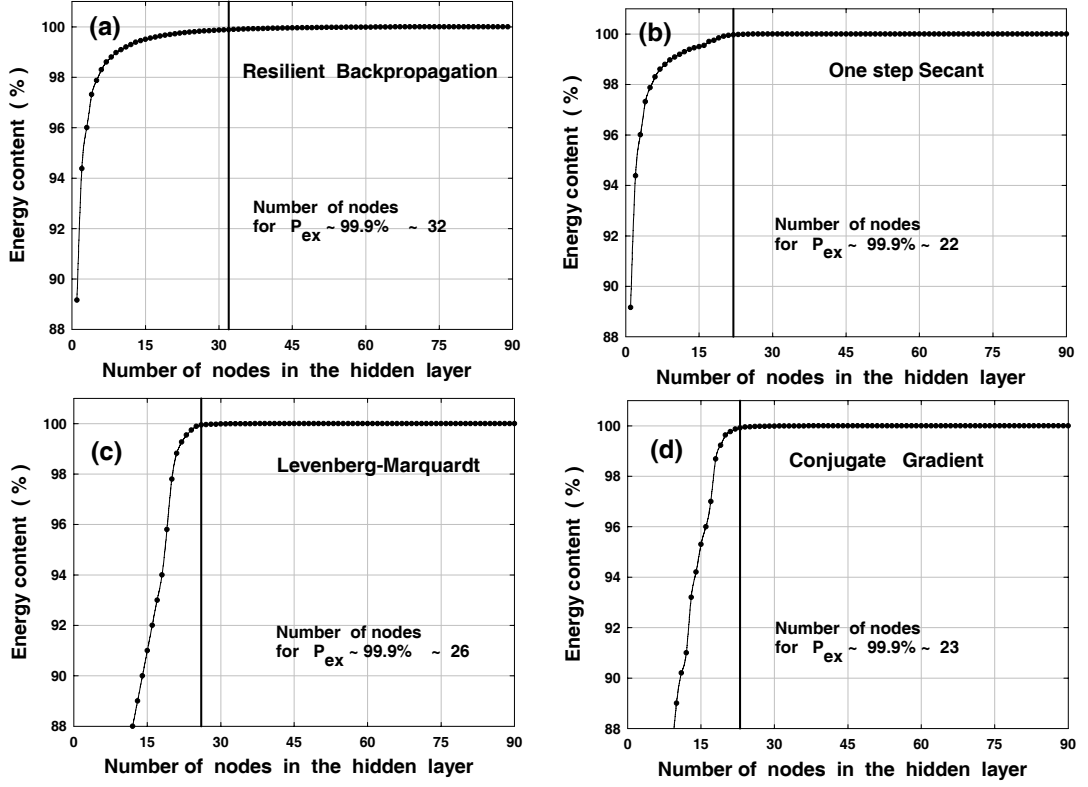


Figure 5.2: Percentage of energy explained ( $P_{ex}$ ) as a function of number of nodes in the hidden layer for some representative algorithms : (a) Resilient backpropagation (b) One Step Secant (c) Levenberg-Marquardt algorithm and (d) Conjugate Gradient. The optimum number of nodes for  $P_{ex} \sim 99.9\%$  is also marked in the figures by full vertical lines.

between  $\sim 22$  to  $\sim 32$ . Except for the Backpropagation-Momentum algorithm which requires only  $\sim 5$  nodes, the remaining algorithms are also found to yield optimum performance with  $\sim 20$  to  $\sim 30$  nodes in the hidden layer. The reason for Backprop momentum requiring too few nodes can be understood from the manner in which the algorithm is trained. In this algorithm, momentum term is added to the Backprop to enhance the training time with a slight compromise on the performance of the network. This effect is seen in our case also where we see the Backprop Momentum algorithm yielding the worst MSE value compared to all other algorithms.

On examining Fig 5.2 and Table 5.5 one can arrive at the following conclusions : (i) None of the ANN algorithms used in this work are under trained or over trained if about 35 nodes are used in the hidden layer. (ii) Increasing the number of nodes beyond 35 results in only a marginal reduction in the MSE. (iii) The MSE value

yielded by the Levenberg-Marquardt method with 35 nodes is found to be the lowest compared to all other ANN algorithms. (iv) Increasing the number of nodes from 35 to 90 leads to the problem of overfitting in the Levenberg-Marquardt method. (v) For the remaining algorithms no overfitting problem is seen when 90 nodes are used in the hidden layer. The overfitting of the Levenberg-Marquardt (with 90 nodes in the hidden layer) is most probably related to the way in which the training is performed in this algorithm, more specifically how the algorithm accounts for error as well as the gradient information based on blending between the gradient descent method and the Gauss Newton rule. The Levenberg-Marquardt trains in such a way that large steps are taken in the direction of low curvature to skip past the plateaus quickly, and smaller steps are taken in the direction of high curvature to slowly converge to the global minima. Thus every narrow valley or plateau, even if as a result of noise in the data, is important for this method. Hence, when larger number of nodes are presented ( i.e, 637 weights for the 90 nodes versus 252 weights for the 35 nodes in the hidden layer), the algorithm becomes sensitive even to the noise values present in the data, which with lesser number of nodes could have been ignored. The source of noise in our training/test data-base is as result of inherent fluctuations in the shower development process. On the basis of the above argument one can thus safely use 35 nodes in the hidden layer for all the algorithms.

It is worth mentioning here that the modification of the ANN structure by analyzing how much each node contributes to the actual output of the neural network and dropping the nodes which do not significantly affect the output is also referred to as pruning. The basic principle of pruning relies on the fact that if two hidden nodes give the same outputs for every input vector, then the performance of the neural network will not be affected by removing one of the nodes in the hidden layer. In the SVD approach, redundant hidden nodes cause singularities in the weight matrix which can be identified through inspection of its singular values. A non-zero number of small singular values indicates redundancy in the initial choice for the number of hidden layer nodes and the approach can be safely used for eliminating these nodes to attain the pruned network model.

A plot of the mean square error as a function of the number of nodes in the hidden layer for the most popular standard backpropagation network and the Lavenberg-Marquardt algorithm with Sigmoid transfer function is shown in Fig 5.3a. While the MSE at the end of the training, for 35 nodes in the hidden layer, is  $\sim 0.1032$  for the backpropagation network, the corresponding value for the Lavenberg-Marguardt algorithm is  $\sim 0.0171$ . Although the MSE yielded by the Lavenberg-Marguardt algorithm is found to be lower than the MSE values of other training algorithms, including the backpropagation algorithm, the reason for showing the MSE for the backpropagation algorithm is mainly because it has been considered as a "work-horse" in the field of neural computation.

The variation of the MSE as a function of number of iterations for all ANN algorithms used, is shown in Fig 5.3b. The number of neurons in the hidden layer

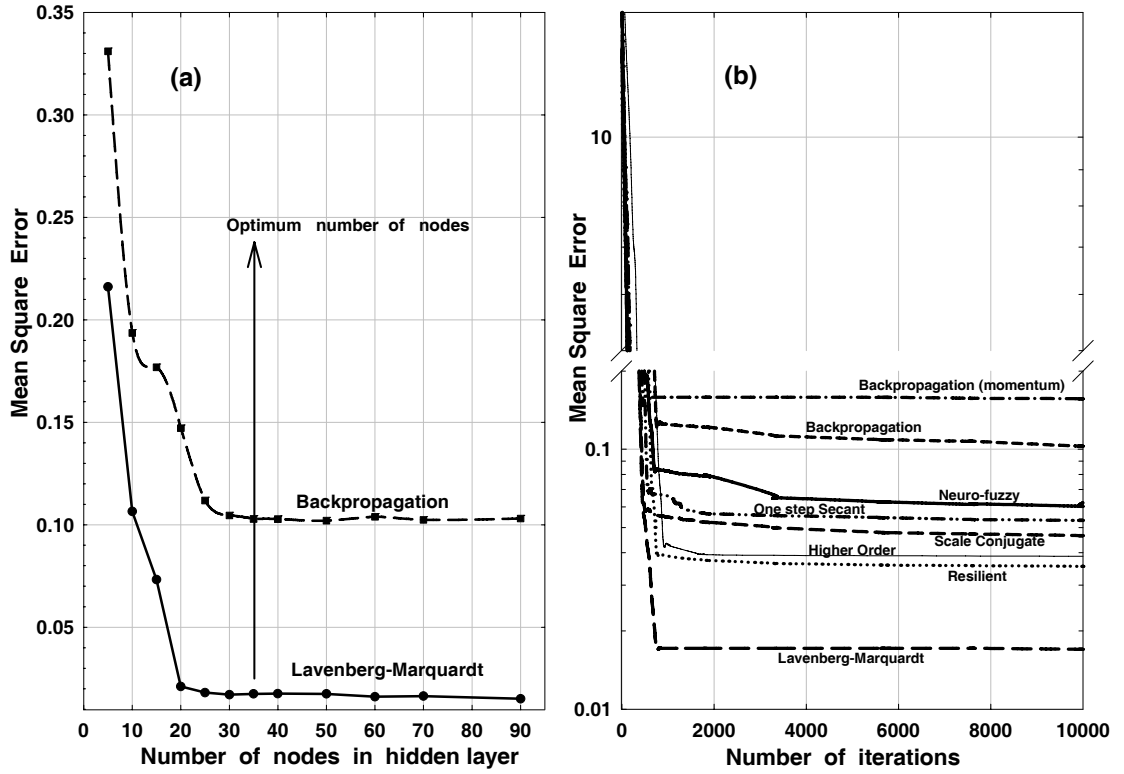


Figure 5.3: (a) Mean square error as a function of number of nodes in the hidden layer for the Backpropagation and the Lavenberg-Marquardt algorithms. (b) Mean square error for various ANN algorithms as a function of number of iterations with 35 nodes in the hidden layer.

was thus fixed at 35 nodes for all these algorithms. In all above algorithms, the training is continued till the MSE error reaches a plateau and does not decrease any further. About 10,000 iterations were generally found to be sufficient to train the ANN on various algorithms. This superior convergence of Lavenberg-Marquardt algorithm over the conventionally used backpropagation algorithm and/or resilient backprop is not totally unexpected and has been demonstrated by us on standard benchmark and regression problems [5.36].

It is worth mentioning here that for studying the performance of the various ANN algorithms we have used BIKAS (BARC-IIT Kanpur ANN Simulator) ANN package [5.36] and MATLAB [5.37,5.38] neural net packages. While as, MATLAB has been used for backpropagation, resilient backpropagation, Scale Conjugate, backprop-momentum, Lavenberg-Marquardt, and One Step Secant algorithms, the BIKAS package has been used for Higher Order Network and Neuro-Fuzzy models.



### 5.5.3 Testing and validating Lavenberg-Marquardt method

Since MSE error returned by the Lavenberg-Marquardt algorithm is lower than the MSE error values of other methods including the backpropagation method, we have used only this algorithm on the test data set for a more descriptive analysis. When the test data-base is presented to the network, instead of yielding the desired output as 0.1 or 0.9, the ANN outputs a range of values between 0.1 to 0.9. The broad distributions around 0.1 and 0.9, returned after testing the prior trained ANN algorithm, instead of the desired 0.1 or 0.9, is on account of the inherent shower to shower fluctuations on an event to event basis even though train and test data is generated in a similar manner. The response of the network (i.e., frequency distribution of the selected events) for the test data sample comprising simulated  $\gamma$ -rays and actual background as a function of the ANN output is shown in Fig 5.4a. The results obtained for the validation data sample are shown in Fig 5.4b. Excellent matching of the results obtained for the test and validation data clearly demonstrates that ANN has indeed "learned" and simply not remembered the classification. It is important to mention here that no cut on  $\alpha$  has been applied to the data presented in these figures.

## 5.6 Determination of optimum ANN cut value

For determining the ANN output cutoff value ( $\eta_{cut}$ ), which will optimize the separation of the two event classes (i.e  $\gamma$ -ray and cosmic-rays), one can maximize either Quality Factor ( $QF$ ) or more adequately, statistical significance ( $N_\sigma$ ). Following their standard definitions [5.15], which are given by :

$$QF = \frac{N_\gamma/N_{\gamma 0}}{\sqrt{N_p/N_{p 0}}} = \frac{f_\gamma}{\sqrt{f_p}} \quad (5.5)$$

$$N_\sigma = \frac{N_\gamma}{\sqrt{N_\gamma + 2N_p}} \quad (5.6)$$

where  $N_\gamma$  and  $N_p$  are the number of  $\gamma$ -rays and hadrons, respectively, after classification;  $N_{\gamma 0}$  and  $N_{p 0}$  are the number of  $\gamma$ -rays and hadrons, respectively, before classifier and  $f_\gamma$  and  $f_p$  are the corresponding acceptances for  $\gamma$ -rays and hadrons. Although many groups have used  $QF$  for optimizing the performance of their classification methods [5.11] , we have optimized the performance of the ANN on the basis of maximizing  $N_\sigma$ . The reason for this is the fact that a high value of  $QF$  can also result from tight cuts which can reduce the  $\gamma$ -ray retention capability of the classification method. Furthermore, maximization of  $N_\sigma$  also ensures that classification procedure is not biased unfavorably towards higher energies. Optimization on the basis of maximizing  $N_\sigma$  has also been followed by other groups [5.10,5.12].

It is worth mentioning here that definition of statistical significance ( $N_\sigma$ ) given above can be only used when  $N_\gamma$  is known beforehand which is possible only with

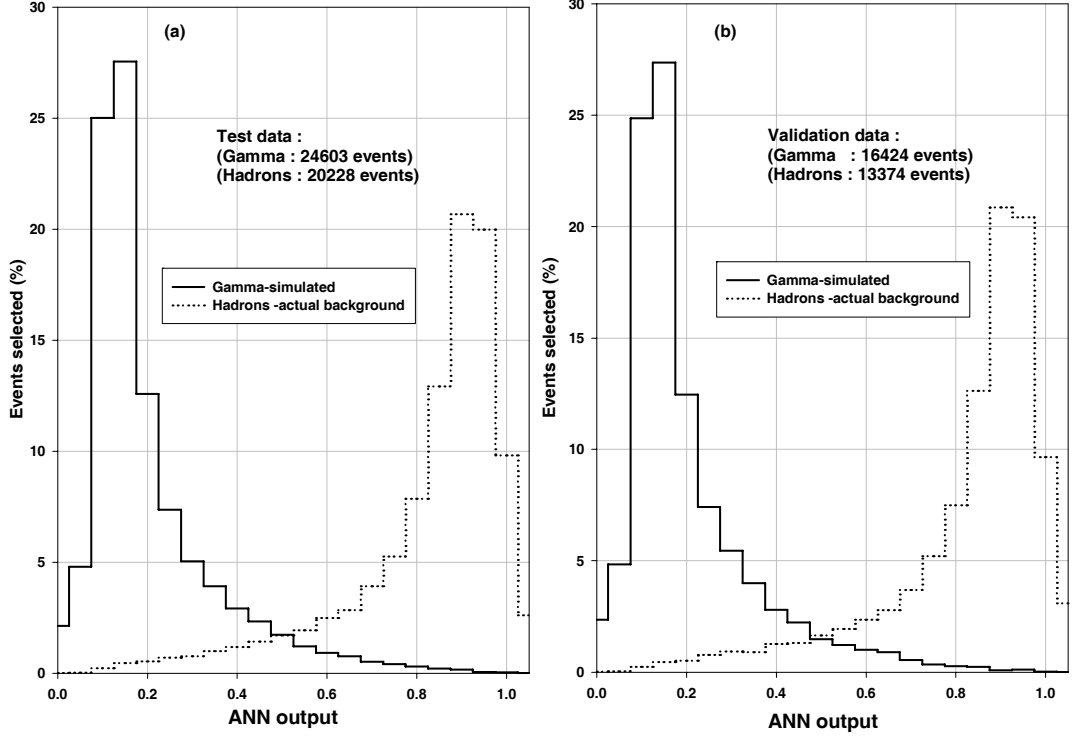


Figure 5.4: (a) ANN output of Lavenberg-Marquardt algorithm in response to simulated  $\gamma$ -rays and actual background events of the test data sample comprising a total 44831 events. (b) Same as (a) except for an independent validation data sample comprising a total 29798 events. No cut on  $\alpha$  has been applied to the data presented in these figures.

simulated data. Since, in case of actual data collected with Cherenkov imaging telescopes,  $N_\gamma$  can also be calculated statistically by subtracting the expected number of background events ( e.g  $27^\circ \leq \alpha \leq 81^\circ$  used by us in [5.26] and in this work) from the  $\gamma$ -ray domain events (e.g  $\alpha \leq 18^\circ$  in our case) the definition of statistical significance given above needs to be modified. While we have used the above expression of  $N_\sigma$  for estimating  $\eta_{cut}$ , the significance of the  $\gamma$ -ray events found in the actual Crab Nebula data has been calculated by following a more rigorous method of using maximum likelihood ratio of Li and Ma [5.39].

The value of  $\eta_{cut}$  defines the decision boundary between the two event species and in order to determine its optimum value we used a data sample of about 12953 events (mixture of 8865 simulated  $\gamma$ -ray and 4088 actual cosmic-ray events). The zenith angle range of these events was again chosen to be in the range  $(0-45)^\circ$ . Since

the value of  $N_\sigma$  depends critically on the number of  $\gamma$ -rays present in the data we have considered only  $N_{\gamma 0} \sim 177$  ( i.e  $\sim 2\%$  of the total  $\gamma$ -rays present in the data sample) for determining the optimum value of  $\eta_{cut}$ . The event is classified as a  $\gamma$ -ray like event only if the corresponding ANN output ( $\eta$ ) is  $\leq \eta_{cut}$  and  $\alpha \leq 18^\circ$ . The calculation was performed by varying  $\eta_{cut}$  from 0.05 to 1.0 in steps of 0.05 and recording  $N_\sigma$  at each value of  $\eta_{cut}$ . The results of this study are given in Fig 5.5, which shows variation of  $N_\sigma$  as a function of  $\eta_{cut}$  for the Levenberg-Marquardt based ANN algorithm. On examining this figure one can see that maximum value

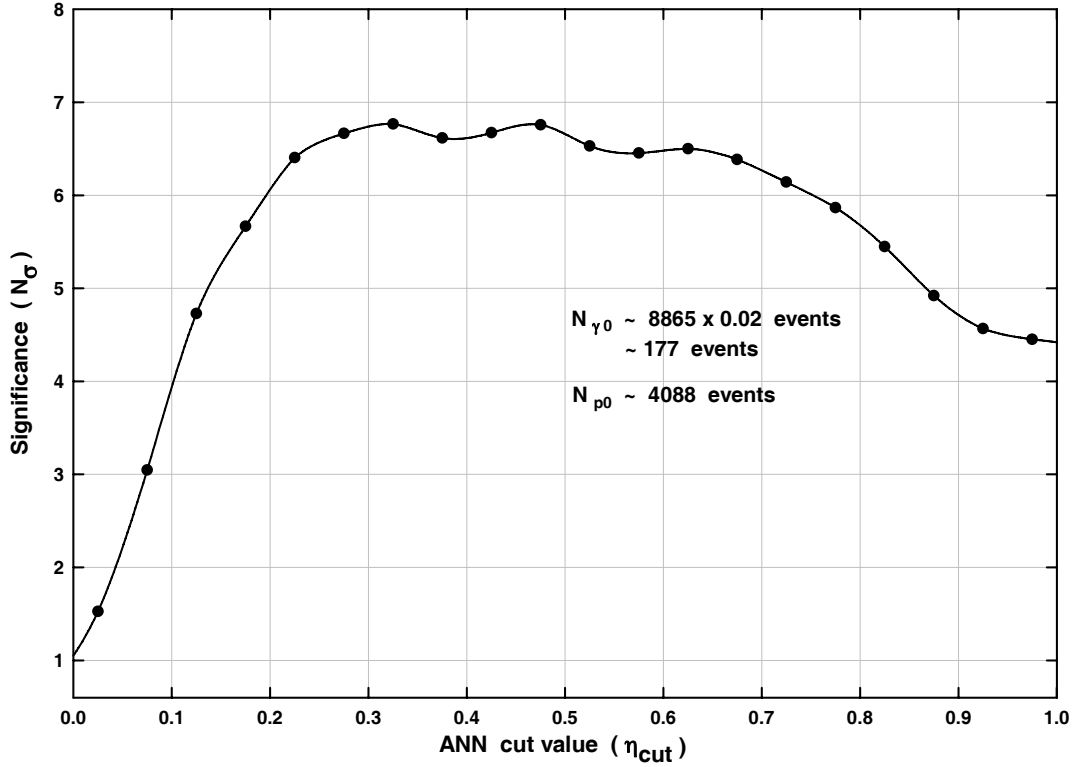


Figure 5.5: Variation of statistical significance ( $N_\sigma$ ) as a function of ANN cut value ( $\eta_{cut}$ ) for the Levenberg-Marquardt algorithm.

of  $N_\sigma \sim 6.8\sigma$  is obtained at  $\eta_{cut} \sim 0.475$ . The above data has also been used for evaluating the performance of other ANN algorithms and finding their optimum  $\eta_{cut}$  values. The results of this study are summarized in Table 5.6 where, in addition to  $N_\sigma$  values yielded by different algorithms, we also give the corresponding  $\eta_{cut}$  range within which  $N_\sigma$  stays constant. The lower value of  $\eta_{cut}$  defines the tight cut and higher value designates the loose cut.

Table 5.6: Maximum value of the statistical significance  $N_\sigma$  yielded by various ANN algorithms along with corresponding  $\eta_{cut}$  range with in which  $N_\sigma$  stays constant. The lower value of  $\eta_{cut}$  defines the tight cut and higher value defines the loose cut.

<b>Algorithm</b>	<b><math>\eta_{cut}</math></b>	<b><math>N_\sigma</math></b>
<i>Backpropagation</i>	0.40 – 0.67	5.21
<i>Backpropagation monemtum</i>	0.30 – 0.57	5.33
<i>Resilient Backprop</i>	0.42 – 0.67	5.25
<i>Scale Conjugate</i>	0.40 – 0.67	4.80
<i>One Step Secant</i>	0.42 – 0.67	5.25
<i>Lavenberg Marquardt</i>	0.30 – 0.62	6.80
<i>Higher Order</i>	0.40 – 0.70	4.80
<i>Neuro Fuzzy</i>	0.40 – 0.67	4.47
<i>Dynamic Super cut</i>	– – –	6.09

The value of  $N_\sigma$  achieved with Dynamic Supercuts is also shown in the table for comparison. It is quite evident from the table that out of 8 different ANN algorithms studied here, Levenberg-Marquardt algorithm yields the best results. The value of  $N_\sigma$  for other algorithms is found to vary from  $\sim 4.5\sigma$  (Higher order network) to  $\sim 5.3\sigma$  (backprop-momentum). Because of the superior performance of the Levenberg-Marquardt algorithm, we will only use this algorithm for analyzing the actual Crab Nebula data.

Referring back to Fig 5.5, since the change in  $N_\sigma$  is insignificant when  $\eta_{cut}$  is varied from  $\sim 0.3$  to  $\sim 0.5$ , we will use a value of  $\eta_{cut} \sim 0.5$  for analyzing the actual Crab Nebula data. Admittedly, using  $\eta_{cut} \sim 0.5$  also increases the cosmic ray background. The reason for choosing the higher  $\eta_{cut}$  value is to ensure that we retain maximum number of  $\gamma$ -rays from the source. For sources which are weaker than the Crab Nebula one can use  $\eta_{cut} \sim 0.3$  so that contamination from more background can be reduced. Since our main preference is to observe relatively stronger sources such as blazars using  $\eta_{cut} \sim 0.5$  is an obvious choice if we want to measure their energy spectra beyond energies of  $\sim 10\text{TeV}$ . Following this approach of choosing the tight cuts for detecting weaker/new sources and loose cuts for obtaining the energy spectrum, is a well known procedure which is adopted by almost all the groups who work on Cherenkov imaging telescopes.

## 5.7 Application of the ANN methodology to the Crab Nebula data collected with the TACTIC telescope

In order to study the  $\gamma$ /hadron segregation potential of the ANN methodology, we have applied this selection method to the Crab Nebula data collected with the TACTIC telescope. For this purpose we reanalyzed the Crab Nebula data for  $\sim 101.44$  h collected during Nov. 10, 2005 - Jan. 30, 2006. The zenith angle during the observations was  $\leq 45^\circ$  and the data was collected with inner 225 pixels ( $\sim 4.5^\circ \times 4.5^\circ$ ) of the full imaging camera with the innermost 121 pixels ( $\sim 3.4^\circ \times 3.4^\circ$ ) participating in the trigger. The standard two-level image 'cleaning' procedure with picture and boundary thresholds of  $6.5\sigma$  and  $3.0\sigma$ , respectively was employed to obtain the clean Cherenkov images. Details of this analysis procedure and the data collecting methodology for this period can be found in [5.27]. The purpose of this image cleaning procedure is to take care of the fluctuations in the image which arise due to electronic noise and night sky background variations. These clean Cherenkov images were then characterized by calculating their standard image parameters like LENGTH, WIDTH, DISTANCE,  $\alpha$ , SIZE and FRAC2. Before investigating the  $\gamma$ /hadron segregation potential using ANN methodology, we will first apply the standard Dynamic Supercuts procedure [5.5] to the data for extracting the  $\gamma$ -ray signal from the background cosmic-ray events.

The cut values used for the analysis are the following :  $0.11^\circ \leq LENGTH \leq (0.260 + 0.0265 \times \ln S)^\circ$ ,  $0.06^\circ \leq WIDTH \leq (0.110 + 0.0120 \times \ln S)^\circ$ ,  $0.52^\circ \leq DISTANCE \leq 1.27^\circ \cos^{0.88} \theta$ ,  $SIZE \geq 450 d.c$  ( where  $6.5$  digital counts  $\equiv 1.0$  pe ),  $\alpha \leq 18^\circ$  and  $FRAC2 \geq 0.35$ . It is important to emphasize here that the Dynamic Supercuts  $\gamma$ -ray selection criteria used in the present analysis are the same which we had used in our previous work [5.26] for developing an ANN-based energy reconstruction procedure for the TACTIC telescope. Since the present work uses the same data-base as well as the same energy reconstruction procedure, we will consider the previous work [5.26] as some sort of benchmark for the present study. Admittedly, there may be a scope for optimizing the previously used Dynamic Supercuts further (e.g by using cuts which depend on both energy and zenith angle), but the results of this study will be presented elsewhere.

A well established procedure to extract the  $\gamma$ -ray signal from the cosmic-ray background using a single imaging telescope is to plot the frequency distribution of  $\alpha$  parameter which is expected to be flat for the isotropic background of cosmic events [5.5]. For  $\gamma$ -rays, coming from a point source, the distribution is expected to show a peak at smaller  $\alpha$  values. Defining  $\alpha \leq 18^\circ$  as the  $\gamma$ -ray domain and  $27^\circ \leq \alpha \leq 81^\circ$  as the background region, the number of  $\gamma$ -ray events is then calculated by subtracting the expected number of background events (calculated on the basis of background region) from the  $\gamma$ -ray domain events. The number of  $\gamma$ -

ray events obtained after applying the above cuts are found to be  $\sim(928 \pm 100)$  with a statistical significance of  $\sim 9.40\sigma$ . The significance of the excess events has been calculated by using the maximum likelihood ratio method of Li & Ma [5.39]. The  $\alpha$ -distribution is given in Fig 5.6a and the corresponding differential energy spectrum of the Crab Nebula shown in Fig 5.6b. The details of the energy estimation for determining the energy of a candidate  $\gamma$ -ray source is presented in chapter 6 of this thesis. The procedure followed uses an ANN to estimate the energy of a  $\gamma$ -ray like event on the basis of its image SIZE, Distance and Zenith angle. The differential photon flux per energy bin has been computed using the formula:

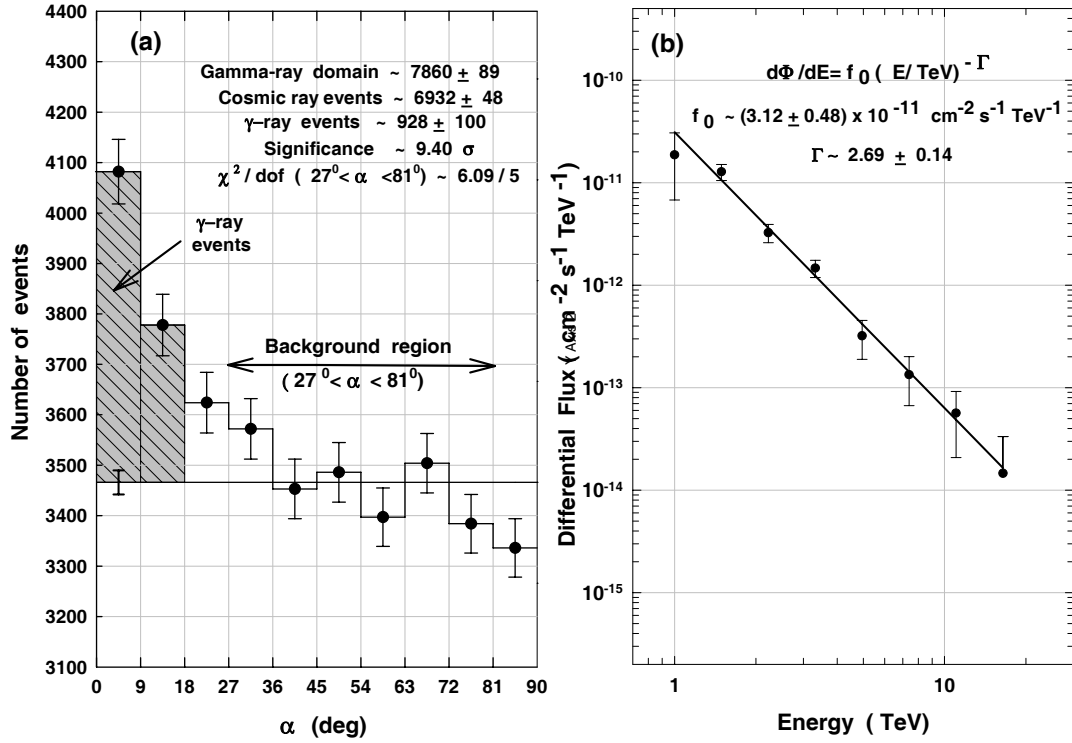


Figure 5.6: (a) Crab Nebula  $\alpha$ -plot for  $\sim 101.44$  h of data using Dynamic Supercuts  $\gamma$ -ray selection criteria. (b) The corresponding differential energy spectrum of the Crab Nebula as measured by the TACTIC telescope.

$$\frac{d\Phi}{dE}(E_i) = \frac{\Delta N_i}{\Delta E_i \sum_{j=1}^5 A_{i,j} \eta_{i,j} T_j} \quad (5.7)$$

where  $\Delta N_i$  and  $d\Phi(E_i)/dE$  are the number of events and the differential flux at

energy  $E_i$ , measured in the  $i$ th energy bin  $\Delta E_i$  and over the zenith angle range of  $0^\circ$ - $45^\circ$ , respectively.  $T_j$  is the observation time in the  $j$ th zenith angle bin with corresponding energy-dependent effective area ( $A_{i,j}$ ) and  $\gamma$ -ray acceptance ( $\eta_{i,j}$ ). The 5 zenith angle bins ( $j=1-5$ ) used are  $0^\circ$ - $10^\circ$ ,  $10^\circ$ - $20^\circ$ ,  $20^\circ$ - $30^\circ$ ,  $30^\circ$ - $40^\circ$  and  $40^\circ$ - $50^\circ$  with effective collection area and  $\gamma$ -ray acceptance values available at  $5^\circ$ ,  $15^\circ$ ,  $25^\circ$ ,  $35^\circ$  and  $45^\circ$ . The number of  $\gamma$ -ray events ( $\Delta N_i$ ) in a particular energy bin is calculated by subtracting the expected number of background events, from the  $\gamma$ -ray domain events. The  $\gamma$ -ray differential spectrum, shown in Fig 5.6b, has been obtained after using appropriate values of effective collection area and  $\gamma$ -ray acceptance efficiency (along with their energy and zenith angle dependence). A power law fit ( $d\Phi/dE = f_0 E^{-\Gamma}$ ) with  $f_0 \sim (3.12 \pm 0.48) \times 10^{-11} \text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$  and  $\Gamma \sim 2.69 \pm 0.14$  is also shown in Fig 5.6b. The fit has a  $\chi^2/dof \sim 3.64/6$  with a corresponding probability of  $\sim 0.72$ . Details of the energy reconstruction procedure can be seen in [5.31] which uses 3:30:1 ANN configuration with SIZE, DISTANCE and Zenith angle as the inputs to the neural net.

While applying the already trained Lavenberg-Marquardt based ANN network, with 6:35:1 configuration, for extracting the  $\gamma$ -ray signal from the data, the number of  $\gamma$ -ray events are found out to be  $\sim (1141 \pm 106)$  with a statistical significance of  $\sim 11.07\sigma$ . A value of  $\eta_{cut} \sim 0.50$  has been used for selecting  $\gamma$ -ray events and only those events are allowed to go for classification with ANN, which satisfy the pre-filtering cuts ( $\text{SIZE} \geq 50\text{pe}$  and  $0.4^\circ \leq \text{DISTANCE} \leq 1.35^\circ$ ). The  $\alpha$ -distribution of the ANN selected events is given in Fig 5.7a, while as the corresponding differential energy spectrum is shown in Fig 5.7b. A power law fit ( $d\Phi/dE = f_0 E^{-\Gamma}$ ) with  $f_0 \sim (1.16 \pm 0.14) \times 10^{-11} \text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$  and  $\gamma \sim 2.52 \pm 0.12$  is also shown in Fig 5.7b. The fit has a  $\chi^2/dof \sim 4.58/7$  with a corresponding probability of  $\sim 0.71$ . Reasonably good matching of the Crab Nebula spectrum with that obtained by the Whipple and HEGRA groups [5.40,5.41] reassures that the procedure followed by us for selecting  $\gamma$ -ray like events as well as obtaining the energy spectrum of a source, is quite reliable.

On comparing the results of Dynamic Supercuts  $\gamma$ -ray selection procedure (Fig 5.6) with the Lavenberg-Marquardt based ANN network (Fig 5.7) it is evident that the performance of the later is somewhat superior, both with regard to improving the statistical significance of the  $\gamma$ -ray signal as well as in selecting more number of  $\gamma$ -rays. Although the improvement (i.e gain of  $\sim 213$   $\gamma$ -ray like events along with signal enhancement from  $9.4\sigma$  to  $11.07\sigma$ ) looks to be only modest, the main advantage accruing from the ANN methodology is that it is more efficient at higher energies which has allowed us to extend the Crab Nebula energy spectrum up to an energy of  $\sim 24\text{TeV}$ . At  $\gamma$ -ray energies above  $\sim 9\text{TeV}$ , the Lavenberg-Marquardt based ANN network selects  $\sim (85 \pm 28)$  events as against  $\sim (24 \pm 9)$  events selected by the Dynamic Supercuts procedure.

When a value of  $\eta_{cut} \sim 0.30$  is used, the number of  $\gamma$ -ray events are found out to be  $\sim (680 \pm 67)$  with a statistical significance of  $\sim 10.49\sigma$  and this is in perfect

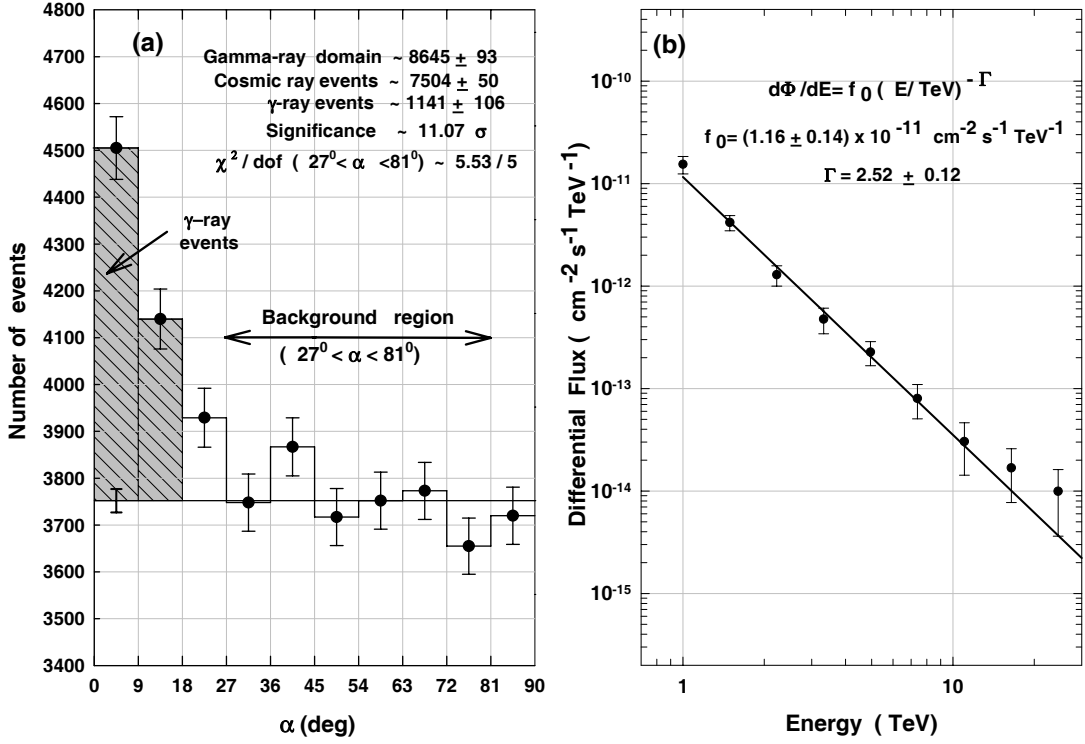


Figure 5.7: (a) Crab Nebula  $\alpha$ -plot for  $\sim 101.44$  h of data using Lavenberg-Marquardt based ANN network  $\gamma$ -ray selection criteria. (b) The corresponding differential energy spectrum of the Crab Nebula when ANN network is used for selecting  $\gamma$ -ray like events.

agreement with the discussion presented in Section [5.6] above. Although the use of tight cut (i.e  $\eta_{cut} \sim 0.3$ ) yields almost same statistical significance (ignoring slight degradation) as compared to  $\eta_{cut} \sim 0.5$  cut case, the number of  $\gamma$ -rays retained are significantly less and it is just for this reason that we preferred to use a somewhat loose cut  $\eta_{cut} \sim 0.5$ .

The performance of the Lavenberg-Marquardt based ANN network was further validated by applying it to  $\sim 201.72$  hours of on-source data collected on Mrk 421 with the TACTIC telescope during Dec. 07, 2005 to Apr. 30, 2006. The total data used here also includes observations from Dec. 27, 2005 to Feb. 07, 2006 when the source was found to be in a high state by the TACTIC telescope as compared to the rest of the observation period [5.27]. When already trained ANN is used for extracting the  $\gamma$ -ray signal from the data, the number of  $\gamma$ -ray events are found out to be  $\sim (1493 \pm 121)$  with a statistical significance of  $\sim 12.60\sigma$ . On comparing



these results with that obtained by using Dynamic Supercuts [5.27] which yields  $\sim(1236\pm110)$   $\gamma$ -ray events with a statistical significance of  $\sim11.49\sigma$ , it is reassuring to find that the ANN method is indeed more efficient than the Dynamic Supercuts method. Furthermore, as expected, no signature of a  $\gamma$ -ray signal is seen when the ANN method is applied to  $\sim 29.65$  hours of off-source data. The results obtained with the ANN method ( $\sim 60\pm42$  with a statistical significance of  $\sim1.46\sigma$ ) compare well with the results reported by us earlier using Dynamic Supercuts [5.27] ( $\sim 28\pm20$  with a statistical significance of  $\sim0.71\sigma$ ).

Successful detection of  $\gamma$ -rays from Mrk-421 thus clearly demonstrates the capability of the properly trained ANN to extract a  $\gamma$ -ray signal from a source other than the Crab Nebula. It also indicates that the generalization capability of the ANN can be enhanced if it is trained with the experimental data collected from different directions having somewhat variable sky brightness.

## 5.8 Comparison of Dynamic Supercuts and ANN analysis methods

A detailed study for comparing the performance of Dynamic Supercuts and ANN analysis methods has also been conducted by us so that the overall  $\gamma$ -ray retention capability of the Dynamic Supercuts and ANN analysis methods can be compared. One of the ways to study this is to use the Monte Carlo simulated data for  $\gamma$ -rays and plot the dependence of effective collection areas as a function of primary energy for the two  $\gamma$ -ray selection methodologies. The results of this study are shown in Fig 5.8 where effective collection areas for the two  $\gamma$ -ray selection methodologies are plotted as a function of energy for two representative zenith angle values of  $15^\circ$  and  $35^\circ$ . Apart from showing the effective areas ( i.e  $A_\gamma(E)f_\gamma(E)$ ) for the two  $\gamma$ -ray selection methodologies, the corresponding effective area when no cuts are applied to the data ( i.e  $A_\gamma(E)$ ) is also shown for comparison. The results displayed in the figure clearly indicate that the efficiency of Dynamic Supercuts is biased towards lower energies ( particularly at lower zenith angles). On the other hand, it is the superior performance of Lavenberg-Marquardt based ANN network ( i.e more collection area at higher energies) which has enabled us to retain relatively higher number of events at energies above  $\sim 9$  TeV in the actual data as compared to the Dynamic Supercuts procedure.

The above conclusion has been further validated by obtaining scatter plots of various image parameters and the results of this study are shown in Fig 5.9. This figure displays scatter plots of LENGTH, WIDTH, DISTANCE and FRAC2 as a function SIZE for  $\sim 8358$  events which have been characterized as  $\gamma$ -ray like by the ANN algorithm and have  $\alpha \leq 18^\circ$ . For comparison, the Dynamic Supercuts boundaries are also shown in the figure as full lines. It is quite evident from the figure that the ANN method is not just selecting the same population of events

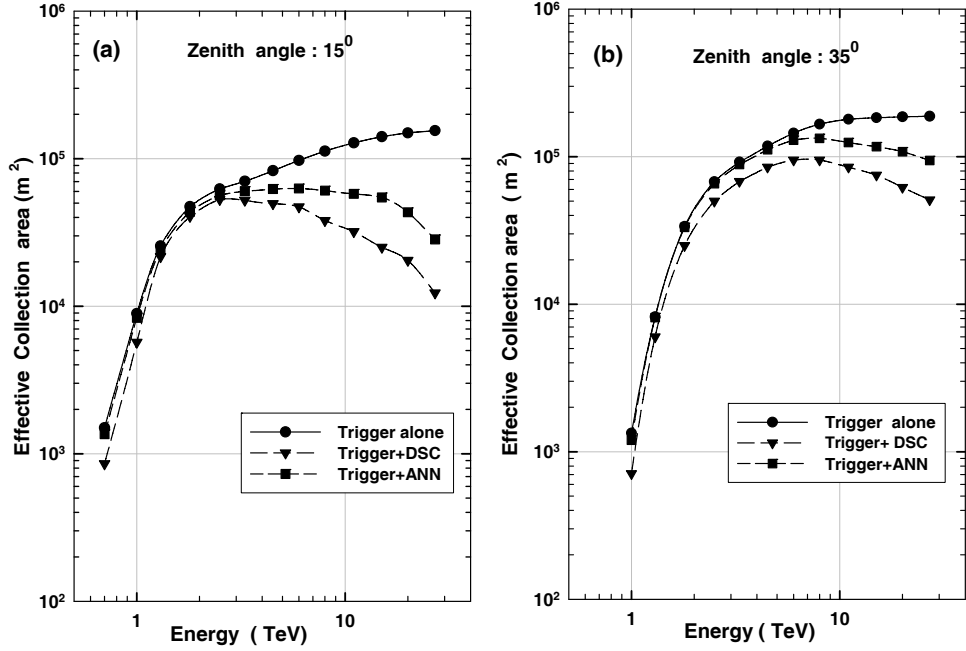


Figure 5.8: Effective collection area as a function of the primary  $\gamma$ -ray energy for simulated  $\gamma$ -ray at showers zenith angles of (a)  $15^\circ$  and (b)  $35^\circ$ . While top most curve (labeled as Trigger alone) gives the effective area when no cuts are applied to the data, the remaining 2 curves (labelled as Trigger+ DSC and Trigger+ANN) represent when Dynamic Supercuts and ANN analysis methods, respectively are applied to the data.

as the Dynamic Supercuts but the ANN is also sensitive to selecting events which lie outside the strict Dynamic Supercuts boundaries. An alternative way to assess the residual population of events selected by ANN is to perform a logical NOT selection between the ANN and the Dynamic Supercuts methods. On performing this selection the number of  $\gamma$ -ray events are found out to be  $\sim(453 \pm 74)$  with a statistical significance of  $\sim 6.27\sigma$  which again suggests that the ANN method is more useful than the Dynamic Supercuts methods while determining the energy spectrum of a  $\gamma$ -ray source. On performing a logical AND selection between the ANN and the Dynamic Supercuts methods the number of  $\gamma$ -ray events yielded are  $\sim(655 \pm 71)$  corresponding to a statistical significance of  $\sim 9.50\sigma$ .

In order to understand the performance of ANN for  $\gamma$ -rays at higher energies (i.e, the events which eventually contribute to the last 3 energy bins of Fig 5.9) Fig 5.10, displays the scatter plot of  $\sim 606$  events which have been characterized as  $\gamma$ -ray like

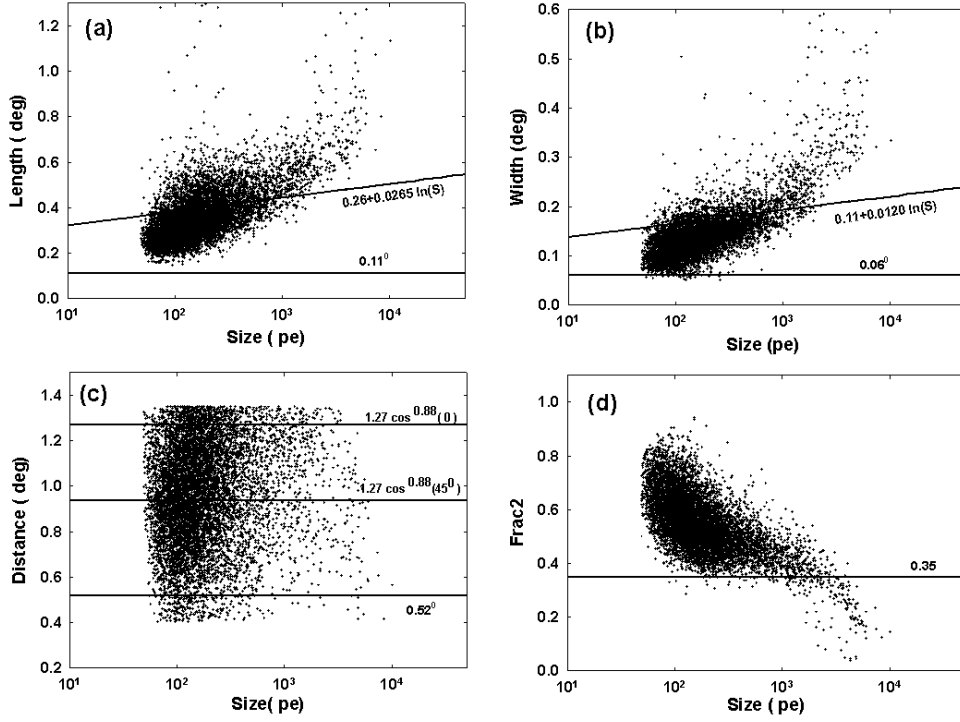


Figure 5.9: Scatter plots of (a) LENGTH (b) WIDTH (c) DISTANCE and (d) FRAC2 as a function of SIZE which have been characterized as  $\gamma$ -ray like by the ANN and have  $\alpha \leq 18^\circ$ . The Dynamic Supercuts boundaries are also shown in the figure as full lines.

by the ANN and which have their  $\alpha \leq 18^\circ$ . In other words the data presented in this figure represents a subsample of the data used in Fig 5.9 with an additional condition that the  $\gamma$ -ray like events should have energies above  $\sim 9$  TeV. The capability of the ANN in selecting events which lie outside the strict Dynamic Supercuts boundaries is again evident from the figure. For example, presence of relatively large number of event outside the LENGTH cut boundary (Fig 5.10a) clearly demonstrates that the efficiency of Dynamic Supercuts in retaining  $\gamma$ -rays is biased towards lower energies. It is important to point out here that there are background cosmic-ray events also present in Fig 5.9 and Fig 5.10 which are classified as  $\gamma$ -ray like events by the event selection methodology. Since subtraction of the background events (estimated from  $27^\circ \leq \alpha \leq 81^\circ$  region), from the  $\gamma$ -ray domain (defined as  $\alpha \leq 18^\circ$ ), will cancel out these events (in statistical sense) and it does not matter how the energy estimate for background event was obtained.

Since differences in the observed energy spectrum of several active galactic nuclei, especially at higher energies, can be used to study absorption effects at the

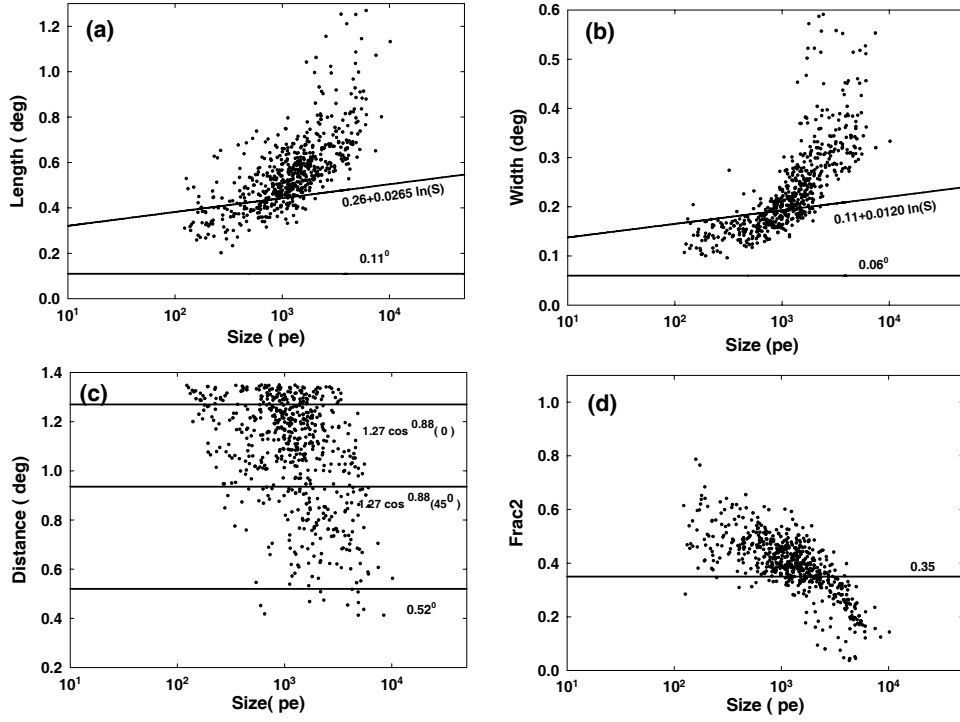


Figure 5.10: Scatter plots of (a) LENGTH (b) WIDTH (c) DISTANCE and (d) FRAC2 as a function of SIZE which have been characterized as  $\gamma$ -ray like by the ANN. Apart from having  $\alpha \leq 18^\circ$  these events also have energy above  $\sim 9$  TeV. The Dynamic Supercuts boundaries are also shown in the figure as full lines.

source or in the intergalactic medium due to interaction of  $\gamma$ -rays with the extragalactic background photons [5.42, 5.43], unarguably, efficient retention of high energy  $\gamma$ -ray events is always preferable. Superior performance of the ANN at higher energies can thus play an important role in understanding the absorption effects at the source or in the intergalactic medium.

It is worth mentioning here that once satisfactory training of the ANN is achieved, the corresponding ANN generated weight-file can be easily used by an appropriate subroutine of the main data analysis program for selecting  $\gamma$ -ray like events. Use of a dedicated ANN software package is thus necessary only during the training of the ANN and is not needed there after. Also, compared to the conventional  $\gamma$ /hadron separation methods, the ANN-based procedure also offers advantages like applicability over a wider zenith angle range and implementation ease.

## 5.9 Summary

Atmospheric Cherenkov imaging telescopes, especially Monoscopic systems, have to cope up with a deluge of cosmic-ray background events and the capability to suppress these against the genuine  $\gamma$ -rays is one of the main challenges which limits the sensitivity of these telescopes. The main purpose of this paper is to study the  $\gamma$  / hadron segregation potential of various ANN algorithms for the TACTIC telescope, by applying them to the Monte Carlo simulated and the observation data on the Crab Nebula. The results of our study indicate that the performance of Levenberg-Marquardt based ANN algorithm is somewhat superior to the Dynamic Supercuts procedure especially beyond  $\gamma$ -ray energies of  $\geq 9$  TeV. Since for real world problems it is not an easy task to identify the most suitable ANN algorithm by just having a look at the problem, our results suggest that while investigating the comparative performance of other ANN algorithm, the Levenberg-Marquardt algorithm deserves a serious consideration. The main advantage of using the ANN methodology for  $\gamma$ / hadron segregation work is that it is more efficient in retaining higher energy  $\gamma$ -ray events and this has allowed us to extend the TACTIC observed energy spectrum of the Crab Nebula up to an energy of  $\sim 24$  TeV. Reasonably good matching of the Crab Nebula spectrum as measured by the TACTIC telescope with that obtained by the other groups reassures that the ANN-based  $\gamma$ /hadron segregation method and also the procedure for obtaining the energy spectrum of a  $\gamma$ -ray source are quite reliable.

---

---

## 5.10 References

- [5.1] T.C. Weekes et al., *Astrophys. J.*, 342 (1989) 379
- [5.2] A. Daum et al., *Astropart. Phys.*, 8 (1997) 1.
- [5.3] A.M. Hillas, *Proc. 19th ICRC* , La Jolla, 3 (1985) 445.
- [5.4] F.A. Aharonian et al., *Nucl. Instr. and Meth. A*, 302 (1991) 522.
- [5.5] G. Mohanty et al., *Astropart. Phys.*, 9 (1998) 15.
- [5.6] P.T. Reynolds et al., *J.Phys. G : Nucl. Part. Phys.*, 19 (1993) 1217.
- [5.7] S. Danaher et al., *Astropart. Phys.*, 1 (1993) 357.
- [5.8] A. Haungs et al., *Astropart. Phys* 12 (1999) 145.
- [5.9] A.K. Razdan et al., *Astropart. Phys* 17 (2002) 497.

- [5.10] A. Chilingarian, Pattern Recognition Letters 16 (1995) 33.
- [5.11] P.T. Reynolds and D.J.Fegan, Astropart. Phys., 3 (1995)137.
- [5.12] R.K. Bock et al., Nucl. Instrum. Meth. A516 (2004) 511.
- [5.13] K. Hornik et al., Neural Networks , 2, (1989) 359.
- [5.14] H. Gish, Proc. IEEE Int. Conf. Acoustic Speech Signal Processing, (1990) 1361.
- [5.15] M.D. Richard and L Lippmann, Neural Computation, 3 (1991) 461.
- [5.16] R.A. Vaze et al., Phy. Rev. D., 45(1992)356.
- [5.17] M.J. Lang, J.Phys. G : Nucl. Part. Phys., 24 (1998) 2279.
- [5.18] B.M. Schafer et al., Nucl. Instrum. Meth. A 465 (2001) 394.
- [5.19] V.K. Dhar et al., Meas. Sci. Technol. 21 (2010) 015112.
- [5.20] V.K. Dhar et al., Experimental Astronomy 10 (2000) 487.
- [5.21] S. Bussino and S.M.Mari, Astropart. Phys., 15 (2001) 65.
- [5.22] G.M. Maneva et al., Nucl. Instrum. Meth. A 502 (2003) 789.
- [5.23] D. Dumora et al., Nucl. Phys. B ( Proc. Suppl.), 97 (2001) 255.
- [5.24] De-Mitri et al., Nucl. Instr. and Meth. A. 525( 2004) 132.
- [5.25] S. Westerhoff et al., Astropart. Phys 4 (1995) 119.
- [5.26] V.K. Dhar et al., Nucl. Instr. and Meth. A, 606 (2009) 795.
- [5.27] K.K. Yadav et al., Astropart. Phys., 27 (2007) 447.
- [5.28] R.J. Barlow Guide to use of statistical methods in Physical Sciences, John Wiley and sons 1987.
- [5.29] W.H. Press et al., Numerical Recipies in C, Cambridge University Press India 2007.
- [5.30] P. Boinee et al., Trans. on Engg. Computing and Tech. 7 (2005) ISSN 1305-5313.
- [5.31] J.M. Zurada, Introduction to Artificial Neural Sustems, Jaico Publishing House, Mumbai, 2006.

- [5.32] P.P. Kanjilal et al., Electronic Letters 29 (1993) 17.
- [5.33] D.C. Psychogios et al., IEEE Trans. Neural Networks 5 (1994) 513.
- [5.34] E.J. Teoh, et al., IEEE Trans. Neural Networks 17 (2006) 1623.
- [5.35] S. Chakroborty et al., Speech Communications 52 (2010) 693.
- [5.36] V.K. Dhar et al., Pramana, Jour. of Physics 74, (2010) 307.
- [5.37] <http://www.mathworks.in>.
- [5.38] S.N. Sivananda, Introduction to Neural Networks using MATLAB, McGraw-Hill Companies 2011
- [5.39] T.P.Li and Y.Q. Ma , Astrophys. J, 272 (1983) 317.
- [5.40] A.M. Hillas et al., Astrophys. J, 503 (1998) 744.
- [5.41] F.A. Aharonian et al., Astrophys. J, 614 (2004) 897
- [5.42] F.A. Aharonian et al., Nature. 440 (2006) 1018.
- [5.43] J. Albert et al., Science. 320 (2008) 1752.

# Chapter 6

## Energy reconstruction using ANN

### 6.1 Introduction

Imaging atmospheric Cherenkov telescopes (IACTs) represent the prime instruments for  $\gamma$ -ray astronomy in the TeV energy range [6.1] and [6.2]. With the number of sources established as TeV  $\gamma$ -ray emitters in IACT observations, emphasis is starting to shift from the detection of sources to the precise determination of their  $\gamma$ -ray spectra. It is therefore important to improve the energy resolution of IACTs. The energy of  $\gamma$ -rays is determined from the intensity of IACT images, taking into account the radial distribution of Cherenkov light within the light pool. The study of the spectral energy distributions can yield valuable information about the underlying  $\gamma$ -ray production mechanisms and unusual astrophysical environment characterizing these sources. In addition, differences in the observed energy spectrum of several active galactic nuclei can also be used to study absorption effects at the source or in the intergalactic medium due to the interaction of  $\gamma$ -rays with the extragalactic background photons [6.3,6.4].

Determining the energy of primary  $\gamma$ -rays is an important advantage which endows the atmospheric Cherenkov technique with calorimetric capability. While the light intensity in an image (also known as image SIZE), represents a key parameter for determining the energy of the primary  $\gamma$ -ray, one also has to consider its dependence on the core-position and zenith angle for improving the energy resolution. Since the precise information of core distance is not available with a single imaging telescope, the energy resolution of these telescopes is generally limited to  $\sim 25\text{-}35\%$  [6.5-6.7]. On the other hand, a stereoscopic system allows unambiguous reconstruction of the shower geometry including a direct measurement of core distance which leads to a significant increase in sensitivity and energy resolution of these systems [6.8,6.9].



### 6.1.1 Factors governing the energy reconstruction

- **Statistical fluctuations in the image itself**

For a single telescope, the statistical fluctuations in the number of photoelectrons produced by a typical shower is a limiting factor for reconstruction of the event energy. Additional fluctuations arise from the amplification process in the photomultiplier and from night-sky background contaminating the image.

- **Image truncation**

In order to reduce the influence of the night-sky background, the image intensity is usually summed only over image pixels above a minimum intensity, cutting away the tail of the image. The sum over image pixel amplitudes provides the so-called size parameter used to derive the shower energy. Such a tail cut introduces both additional noise as well as systematic nonlinearities. Thus for low-intensity images a larger fraction of the image is cut compared to the intense images. An additional truncation occurs for images which extend beyond the edge of the camera.

- **Threshold effects**

In the region near the trigger threshold the image intensity detected in the camera will be strongly biased, since showers with upward fluctuations in the image size will have a larger probability of triggering. In the sub-threshold energy region, the mean intensity of triggered images will approach a constant, independent of the shower energy, making the energy determination a difficult task.

- **Fluctuations in the shower development**

Variations in the shower development provide a significant contribution to the energy resolution; particularly relevant are fluctuations in the height of the shower maximum, related primarily to the fluctuation in the depth of the first interaction. Showers with their maximum deeper in the atmosphere have a higher intensity of light within their light pool, both because of the smaller distance between the telescope and the light source, and because of the lower Cherenkov threshold at reduced height.

- **Systematic errors**

All techniques for energy determination rely heavily on Monte Carlo simulations to provide the relation between image parameters and shower energy and to describe the performance of the telescope hardware. Imperfections in the simulations of the air shower, or of the telescopes, or alignment errors and calibration errors not included in the simulations may have a detrimental effect on the energy resolution. Great care must be taken to ensure that the simulations properly reproduce all relevant aspects of the data.

- **Monte Carlo statistics**

Algorithms for energy reconstruction frequently use multi-dimensional lookup tables to convert values of image parameters into energy estimates. Given the time-consuming generation of Monte Carlo events in particular at the higher energies, their number is frequently similar to, or even less than the number of showers detected in the experiment. Statistical errors in these lookup table values can be significant. They can be alleviated either by an efficient choice of variables or by fitting a smooth analytical function.

The main aim of the present work is to use an ANN based procedure for estimating the energy of  $\gamma$ -ray like events, recorded by a single imaging telescope (TACTIC), on the basis of its image SIZE, DISTANCE and zenith angle. The feasibility of employing ANN for energy reconstruction has also been applied for some other applications which are however not related to Cherenkov imaging. While the Wizard collaboration has used it for GILDA imaging silicon calorimeter [6.10], the ANN-based approach has also been used for reconstruction of the energy deposited in the calorimetry system of the CMS detector [6.11] and the hadronic calorimeter of ATLAS, Tilecal [6.12].

Application of ANN to atmospheric Cherenkov imaging data, for distinguishing between  $\gamma$ -ray and cosmic-ray generated Cherenkov events, has been studied by several workers [6.13-6.16]. Promising results have also been reported for the wave-front sampling telescope CELESTE [6.17] where the ANN method was used for not only discriminating  $\gamma$ -ray and cosmic-ray generated Cherenkov events but also for determining the primary energy and the location of the shower core. A detailed case study comparing different multivariate classification methods (classification trees, kernel and nearest-neighbour methods, linear discriminant analysis, support vector machines, neural network etc.) has also been performed in [6.18] using Monte Carlo simulated data generated for the MAGIC telescope. Keeping in view the encouraging results reported in above cited literature, the main thrust of this work is to use ANN for determining the energy of the  $\gamma$ -rays detected by an atmospheric Cherenkov imaging telescope and its comparison with conventional methods. In addition, the other aspects which have been demonstrated are :(i) achieve a lower normalized rms error, than reported earlier [6.19-6.20] (ii) check the interpolation capability of the proposed ANN method with an independent data sample. Finally, the performance of the ANN-based energy reconstruction is validated by revisiting the energy spectrum of the Crab Nebula in the energy range 1-16 TeV as measured by the TACTIC telescope.

## 6.2 Monte Carlo simulations for energy reconstruction of $\gamma$ -rays

The Monte Carlo simulation data used for developing a procedure for energy reconstruction of  $\gamma$ -rays are based on the CORSIKA (version 5.6211) air-shower simulation code [6.21]. The simulated data-base for  $\gamma$ -ray showers used in the present work is the same as has been mentioned in chapter 2, where about 34,000 showers in the energy range 0.2-20 TeV with an impact parameter of upto 250m have been generated. These showers have been generated at 5 different zenith angles ( $\theta = 5^\circ, 15^\circ, 25^\circ, 35^\circ$  and  $45^\circ$ ). Wavelength dependent atmospheric absorption, the spectral response of the PMTs and the reflection coefficient of mirror facets and light cones has also been taken into account while performing the simulations. The number of photoelectrons registered by each pixel has then been subjected to noise injection, trigger condition check and image cleaning. The clean Cherenkov images were characterized by calculating their standard image parameters like LENGTH, WIDTH, DISTANCE, ALPHA, SIZE and FRAC2 [6.22-6.23]. The same simulation data base has also been used, as per the well known standard procedure, for calculating the effective area of  $\gamma$ -rays as a function of energy and zenith angle and, also the  $\gamma$ -ray retention factors when Dynamic Supercuts are applied to the simulated data. Both these inputs are required for determining the energy spectrum of a source once a statistically significant  $\gamma$ -ray signal is observed in the data.

Keeping in view the fact that the Cherenkov light emitted from the electromagnetic cascade is to a first order approximation proportional to the energy of the primary  $\gamma$ -ray, the approach followed in atmospheric Cherenkov imaging telescopes is to determine the energy on the basis of the image SIZE. Since the intensity of the Cherenkov light is a function of core distance, which is not possible to obtain with a single imaging telescope, the angular distance of the image centroid from the camera center (known as the DISTANCE parameter) is generally used as an approximate measure of the impact distance. The energy reconstruction procedure with a single imaging telescope thus involves using SIZE and DISTANCE parameters of the Cherenkov event for determining energy of the primary  $\gamma$ -ray. Although the method has been found to work reasonably well over a restricted zenith angle range of  $\leq 30^\circ$ , there is a need to include zenith angle dependence in the energy reconstruction procedure for allowing data collection over a much wider zenith angle range.

In order to check the performance of various energy reconstruction procedures for the TACTIC telescope, we have divided the simulated data base into two parts so that one part could be used for preparing the data for obtaining parameterized fits (or training the ANN) and the remaining for testing. For smoothening event to event fluctuations, which are inherently present in raw data, we have first calculated  $\langle \text{SIZE} \rangle$  and  $\langle \text{DISTANCE} \rangle$  by clubbing together showers of a particular energy in various core distance bins with each bin having a size of 40m. Furthermore, addi-

tional selection criteria (viz., accepting events with core distance  $>30\text{m}$ ,  $\text{SIZE}>50\text{pe}$  and  $\text{DISTANCE}$  between  $0.4^\circ$  to  $1.4^\circ$ ) has also been used while preprocessing the training data to ensure that the image is robust with minimum possible truncation effects. Imposing a lower bound on the core distance helps in rejecting the events where shower to shower fluctuations in the light intensity are expected to be very large, as most of the light in this region is produced by local penetrating particles whose number can vary quite widely. The final training data file thus consists of a single table with  $\sim 350$  rows. Each row has 4 columns with one column each for energy,  $\langle\text{SIZE}\rangle$ ,  $\langle\text{DISTANCE}\rangle$  and zenith angle. It is worth mentioning here

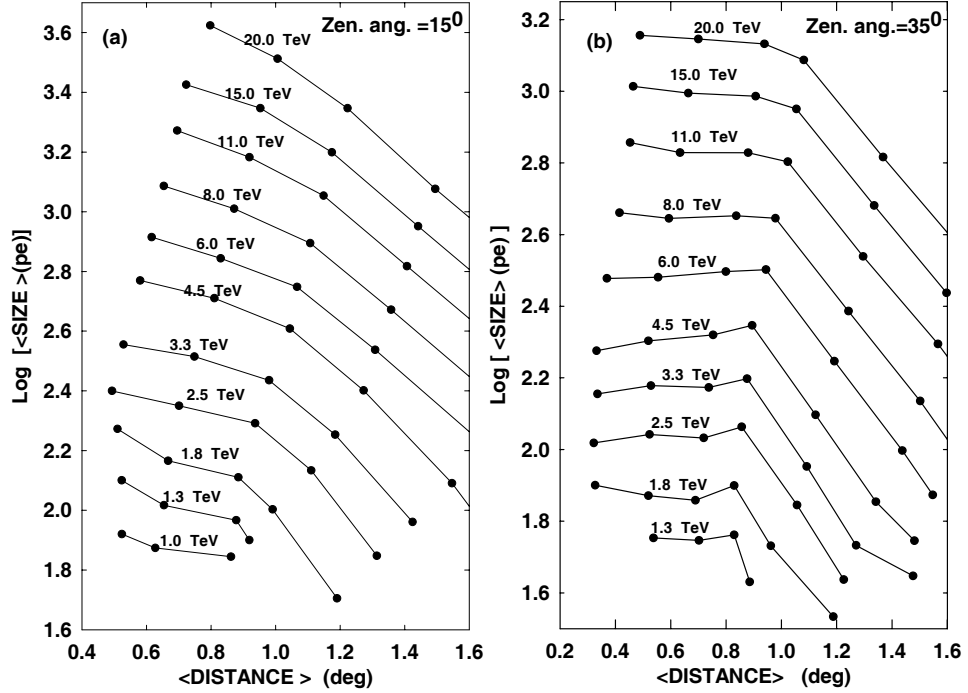


Figure 6.1: Variation of  $\langle\text{SIZE}\rangle$  as a function of  $\langle\text{DISTANCE}\rangle$  for  $\gamma$ -rays from a point source of different energies at zenith angles of (a)  $15^\circ$  and (b)  $35^\circ$ .

that although the raw data used for training is same as used in [6.19], there is a slight difference in the procedure followed for preparing the training data file. Using  $\sim 10,000$  events, a training data file of 350 events was generated by interpolating  $\langle\text{SIZE}\rangle$  at  $\langle\text{DISTANCE}\rangle$  values of  $0.40^\circ$ ,  $0.50^\circ$ , .....  $1.40^\circ$  for each energy and zenith angle. A representative example of the variation of  $\langle\text{SIZE}\rangle$  as a function of  $\langle\text{DISTANCE}\rangle$  for different primary  $\gamma$ -ray energies is shown in Fig 6.1. It is quite evident from this figure that  $\text{SIZE}$  (proportional to the Cherenkov light in an image) is the most important factor which needs to be considered for estimating the

energy of the primary  $\gamma$ -ray. Since, for a fixed  $\gamma$ -ray energy,  $\langle \text{SIZE} \rangle$  also depends on core distance (proportional to DISTANCE parameter of the image for a point  $\gamma$ -ray source) the second factor which needs to be considered is the DISTANCE parameter. On comparing Fig 6.1a and Fig 6.1b, which show the behaviour of  $\langle \text{SIZE} \rangle$  at zenith angles of  $15^\circ$  and  $35^\circ$ , respectively, one finds that the zenith angle dependence cannot be ignored in situations where a wider zenith angle coverage is required.

The performance of a particular energy reconstruction procedure has been evaluated by calculating the relative error in the reconstructed energy ( $\Delta_E$ ), for individual  $\gamma$ -ray events using the test data file. The relative error in the reconstructed energy is defined as  $(E_{\text{estm}} - E_{\text{true}})/E_{\text{true}}$ , where  $E_{\text{true}}$  is the true energy and  $E_{\text{estm}}$  is the estimated energy yielded by the energy reconstruction procedure. The mean value of ( $\Delta_E$ ) as a function of  $E_{\text{true}}$  and energy resolution ( $\sigma(\Delta_E)$ ) defined as the root mean square width of the distribution of  $\Delta_E$  are the main quantities which can be used for comparing the performance of various energy reconstruction procedures. It is worth mentioning that the energy resolution, is sometimes estimated by calculating rms width of the distribution of  $\ln(E_{\text{true}}/E_{\text{estm}})$  [6.5,6.19] or  $\ln(E_{\text{estm}}/E_{\text{true}})$  [6.24]. The only reason for using  $(E_{\text{estm}} - E_{\text{true}})/E_{\text{true}}$  in this work as against  $\ln(E_{\text{estm}}/E_{\text{true}})$  is to follow a more standard and widely accepted definition of energy resolution [6.25]. Nevertheless, it can be easily shown that  $(E_{\text{estm}} - E_{\text{true}})/E_{\text{true}} \sim \ln(E_{\text{estm}}/E_{\text{true}})$ .

## 6.3 Conventional energy reconstruction methods

### 6.3.1 Parameterized fit with DISTANCE and SIZE as variables

The first energy estimation procedure which has been studied here is based on the approach followed by Whipple group [6.5,6.7], where  $\ln(E_{\text{estm}})$  is expressed as a polynomial in  $\ln(\text{SIZE})$  and DISTANCE. The approach assumes that, for a point  $\gamma$ -ray source, DISTANCE parameter of the image provides an approximate measure of the core distance. The validity of this assumption has also been checked for the TACTIC telescope simulation data and the results of the same are presented in Fig 6.2.

The data used in this figure has been generated by clubbing together showers of a particular energy in various core distance bins with each bin having a size of 40m and finding  $\langle \text{DISTANCE} \rangle$  for each core distance bin. Although a strong correlation between core distance and  $\langle \text{DISTANCE} \rangle$  is clearly visible in Fig 6.2, it is worth mentioning here that DISTANCE parameter of a Cherenkov image produced by an individual shower is also dependent on the height of the shower maximum [6.26]. Since for a single telescope it is impossible to determine separately the core distance and height of shower maximum on an event to event basis, obtaining an approximate measure of the core distance on the basis of DISTANCE parameter seems to be the

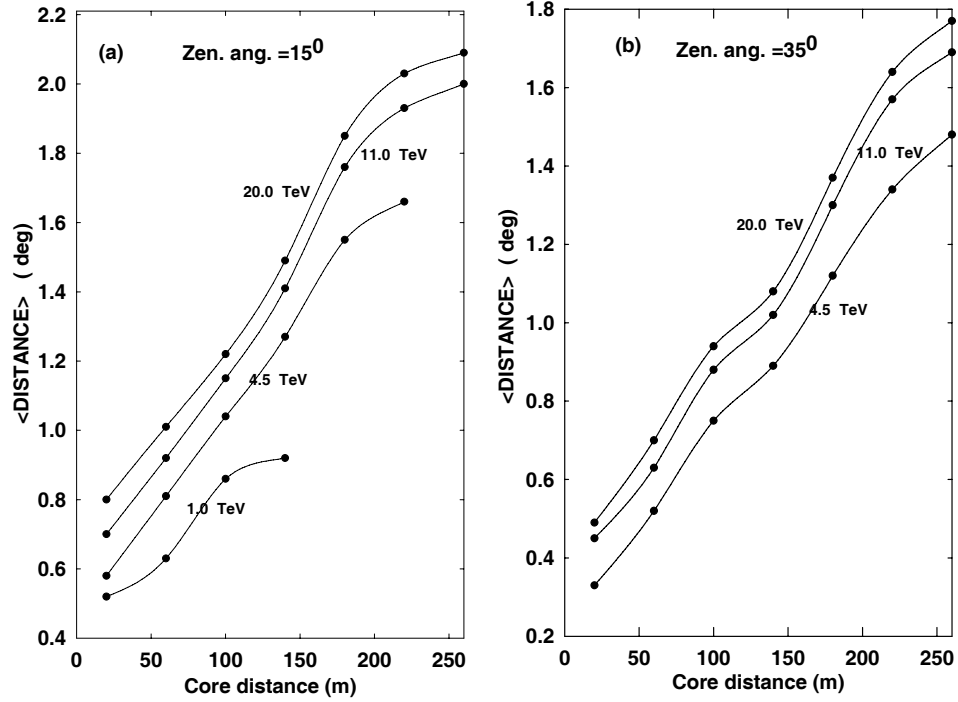


Figure 6.2: Variation of  $\langle \text{DISTANCE} \rangle$  as a function of core distance for  $\gamma$ -rays of various energies from a point source at zenith angles of (a)  $15^\circ$  and (b)  $35^\circ$ .

only viable solution. Ignoring zenith angle dependence and following the Whipple procedure,  $E_{estm}$ , based on image SIZE ( $S$ ) and DISTANCE ( $D$ ) is calculated by using the following relation

$$\ln(E_{estm}) = a_1 + a_2 \ln(S) + a_3 (\ln(S))^2 + a_4(D_0) + a_5(D_0)D \quad (6.1)$$

Choosing  $D_0 = 1.00^\circ$ , the values of  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4(D_0)$  and  $a_5(D_0)$ , obtained after fitting equation (1) to the training data file at zenith angle of  $25^\circ$ , are determined to be the following:  $a_1 \sim -2.8820$ ,  $a_2 \sim 0.7221$ ,  $a_3 \sim 0.0035$ ,  $a_4(D \leq D_0) \sim -0.2005$ ,  $a_5(D \leq D_0) \sim 0.2395$ ,  $a_4(D > D_0) \sim -1.6766$  and  $a_5(D > D_0) \sim 1.7290$ . While the first 3 terms in the above equation use the fact that total intensity of an image is roughly proportional to the energy of the primary, the remaining 2 terms modify this relationship by including the dependence on the core distance also. A plot of relative error in the energy reconstruction obtained for test data sample at zenith angle  $25^\circ$  is shown in Fig 6.3a.

The corresponding relative error in the reconstructed energy ( $\Delta_E$ ) for zenith angle of  $15^\circ$  and  $35^\circ$  is also shown in (b) and (c) if zenith angle dependence is ignored and fit coefficients obtained at zenith angle of  $25^\circ$  are used as such in the

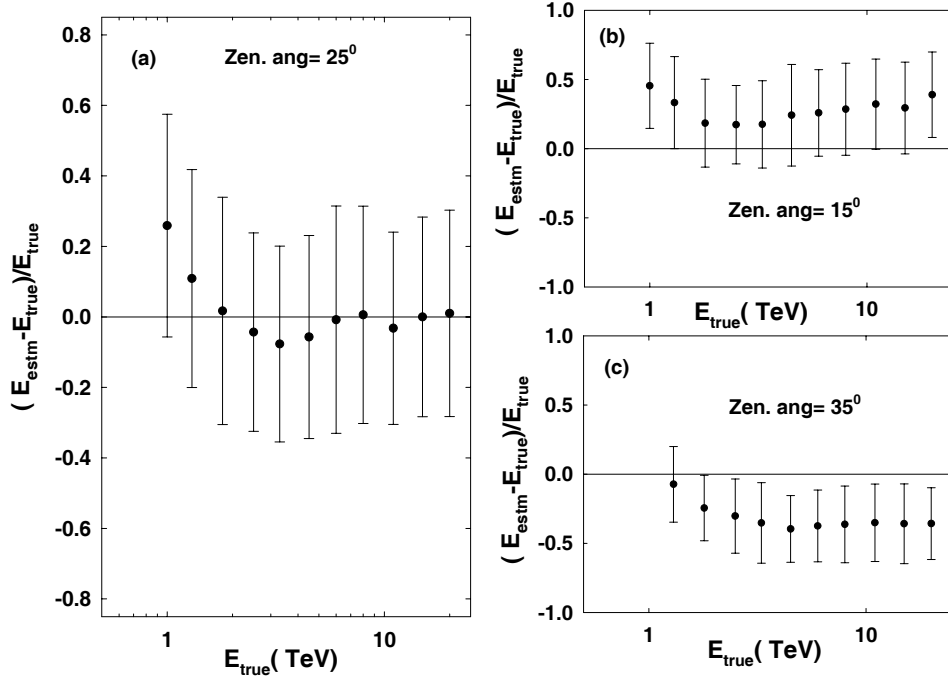


Figure 6.3: (a) Relative mean error in the reconstructed energy ( $\Delta_E = (E_{\text{estim}} - E_{\text{true}})/E_{\text{true}}$ ) as a function of energy for zenith angle of 25° using the energy estimation procedure given by equation (6.1). Relative mean error in the estimated energy as a function of energy for zenith angles of (b) 15° and (c) 35° if zenith angle dependence is ignored in the energy reconstruction procedure.

energy reconstruction procedure at the other two zenith angles also. Although the energy reconstruction procedure yields  $\sigma(\Delta_E) \sim 28\%$  for the data shown in Fig 6.3a, presence of a systematic bias seen in Fig 6.3b and Fig 6.3c ( $\sim 20\%$  and  $\sim -37\%$  at zenith angles of 15° and 35°, respectively) suggests that there is a need to include zenith angle dependence in the energy reconstruction procedure for allowing data collection over a much wider zenith angle range.

### 6.3.2 Parameterized fit with DISTANCE, SIZE and zenith angle as variables

Including the zenith angle ( $z$ ) dependence, in the energy construction procedure, can be in principle implemented by adding one or more zenith angle dependent terms to equation (6.1). The method followed here uses guidance from [6.27] and employs the following relation for estimating the energy.

$$\ln(E_{estm}) = 1.0 + b_1 \ln(S) + b_2 \sqrt{\ln(S)} + b_3 (\ln(S))^2 + b_4 / \cos(z) + b_5 D \quad (6.2)$$

The values of the constants after fitting equation (2) to the training data file at all the 5 zenith angles (i.e 5°, 15°, 25°, 35° and 45°) together are found out to be following :  $b_1 \sim 4.0053$ ,  $b_2 \sim -9.7814$ ,  $b_3 \sim -0.1029$ ,  $b_4 \sim 3.3510$  and  $b_5 \sim 0.7822$ . Plot of relative error in the estimated energy as a function of energy for the test data sample at all 5 zenith angles is shown in Fig 6.4a.

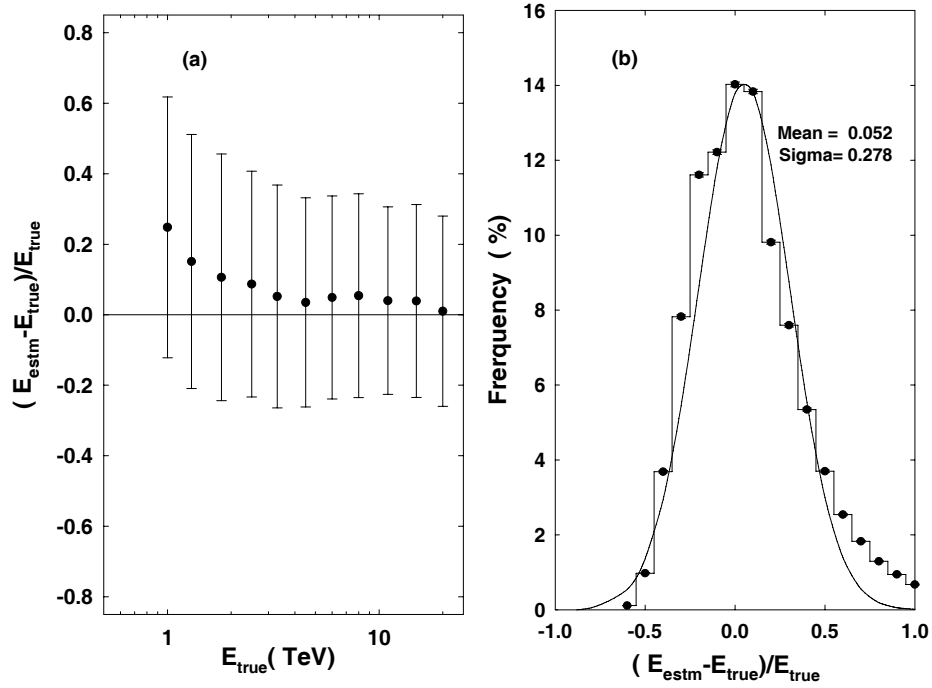


Figure 6.4: (a) Relative mean error in the reconstructed energy ( $\Delta_E = (E_{estm} - E_{true})/E_{true}$ ) as a function of energy for the energy estimation procedure given by equation (6.2). (b) Frequency distribution of  $\Delta_E$  along with a best fit Gaussian distribution to the data.

Frequency distribution of  $\Delta_E$  by considering almost equal number of showers at different energies is shown in Fig 6.4b. The root mean square width of the distribution is  $\sim 31\%$  and the same for the fitted Gaussian distribution is  $\sim 28\%$ . Systematic bias at all energies ( $\sim 7\%$  in the energy range 1.8 TeV to 15.0 TeV) is also seen in Fig.4a which suggests that actual zenith angle dependence in the energy reconstruction procedure is probably more complicated than what has been considered in equation (6.2). An improvement in the energy resolution has been



reported in [6.27] by including an additional parameter called LEAKAGE ( defined as the ratio of light content in the edge pixels to total light content or SIZE) in equation (6.2), to compensate for leakage effects in the relatively small ( $\sim 3^\circ$  diameter ) HEGRA CT1 camera. Since the TACTIC telescope uses a fairly large camera ( $\sim 6^\circ$  diameter) we do not expect the energy resolution to improve if LEAKAGE parameter is also used. However, an attempt to remove the systematic bias was also tried by using a nonlinear model with 2 more zenith angle dependent terms ( viz.,  $D/\cos(z)$  and  $D \ln(S)/\cos(z)$ ) in equation (6.2), but the improvement was found to be only marginal. It is worth mentioning here that while the method of least squares often gives optimal estimates of the unknown parameters, it is very sensitive to the presence of unusual data points in the data used to fit a model. One or two outliers can sometimes seriously skew the results of a least squares analysis.

### 6.3.3 Look-up table method using interpolation in 3 dimensions

The third energy reconstruction method which has been studied here is based on the look-up table method. This method has been used quite extensively by the HEGRA collaboration [6.26]. Although the method was originally developed for the HEGRA stereoscopic array, we essentially follow the same principle here. In this method, we generate the fine grid look-up table by using the training data file. This is done by interpolating the expected  $\langle \text{SIZE} \rangle$  at finer intervals of DISTANCE, energy and zenith angle. The total number of interpolated SIZE values, at a particular zenith angle, comprise  $\sim 4000$  values with DISTANCE parameter ranging from  $0.4^\circ$  to  $1.4^\circ$  and energy values ranging from  $\sim 0.74$  TeV to  $\sim 20$  TeV. In order to perform interpolation in zenith angle, 9 different data files are prepared at  $5^\circ$  interval in the zenith angle range from  $5^\circ$  to  $45^\circ$ . While the above interpolated data has been obtained by fitting polynomial curves of order 3 to the given data points, final energy estimation of an event, on the basis of its SIZE, DISTANCE and zenith angle, uses only linear interpolation. Plot of the relative mean error in the reconstructed energy ( $\Delta_E$ ) as a function of energy for test data sample at all the 5 zenith angles is shown in Fig 6.5a. The frequency distribution of  $\Delta_E$  is shown in Fig 6.5b.

It is quite evident from Fig 6.5a that, barring energy values at 1.0 TeV, 1.3 TeV and 20.0 TeV where  $|\Delta_E|$  is found to be  $>5.0\%$ , the reconstructed energy has a negligible bias at other energy values from 1.8 TeV to 15.0 TeV. The rms width of the frequency distribution is found to be  $\sim 27\%$  and the rms width of the fitted Gaussian distribution is  $\sim 24\%$ . It is important to mention here that a positive bias seen in the relative mean error (Fig 6.3a, Fig 6.4a and Fig 6.5a) at energy values of 1.0 TeV and 1.3 TeV is because of the well known selection effect [6.26] and is due to sub-threshold regime event triggers because of their upward fluctuations in the light yield sometimes. Since events with downward fluctuations in the light yield are unable to trigger the system in this energy range, the energy estimates tend to

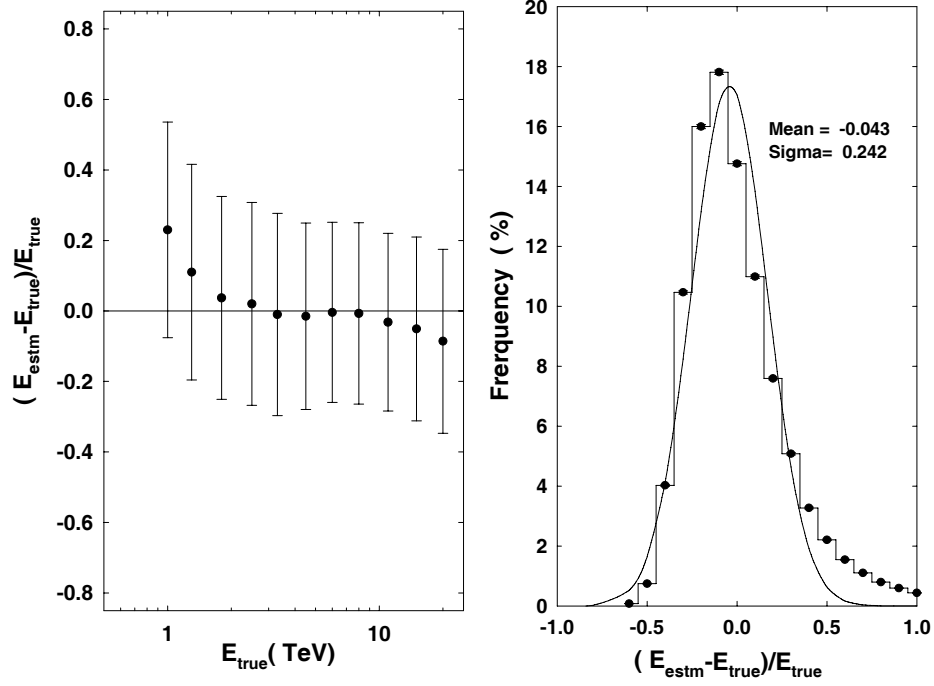


Figure 6.5: (a) Relative mean error in the reconstructed energy ( $\Delta_E = (E_{\text{estm}} - E_{\text{true}})/E_{\text{true}}$ ) as a function of energy for the look-up table based energy estimation procedure. (b) Frequency distribution of  $\Delta_E$  along with a best fit Gaussian distribution to the histogram.

be biased towards larger values.

## 6.4 Energy reconstruction using ANN

### 6.4.1 Training of the network

As already discussed the network requires at least two layers, an input layer and an output layer. In addition, the network can include any number of hidden layers with any number of hidden nodes in each layer. The signal from the input vector propagates through the network layer by layer till the output layer is reached. The output vector represents the predicted output of the ANN and has a node for each variable that is being predicted.

The essence of the training process is to iteratively reduce the error between the predicted value and the target value. While the choice of using a particular error function is problem dependent, there is no well defined rule for choosing the most

suitable error function. We have used the root-mean-squared error (normalized)  $RMS$  [6.28] in this work which is defined as :

$$RMS = \frac{1}{PI} \sqrt{\frac{1}{2} \sum_{p=1}^P \sum_{i=1}^I \left( \frac{D_{pi} - O_{pi}}{D_{pi}} \right)^2} \quad (6.3)$$

where  $D_{pi}$  and  $O_{pi}$  are the desired and the observed values and  $P$  is number of training patterns. The error here depicts the accuracy of the neural network mapping after a number of training cycles have been implemented.

Given the inherent power of Artificial Neural Network (ANN) to effectively handle the multivariate data fitting, we have developed an ANN-based energy estimation procedure for determining the energy of the primary  $\gamma$ -ray on the basis of its image SIZE, DISTANCE and zenith angle. The procedure followed by us uses a 3:30:1 (i.e 3 nodes in the input layer, 30 nodes in hidden layer and 1 node in the output layer) configuration of the ANN with resilient back propagation training algorithm [6.29] to estimate the energy of a  $\gamma$ -ray event on the basis of its image SIZE, DISTANCE and zenith angle. The 3 nodes in the input layer correspond to zenith angle, SIZE and DISTANCE, while the 1 node in the output layer represents the expected energy (in TeV) of the event. As already mentioned earlier, the training data comprises  $\sim 350$  events where  $\langle \text{SIZE} \rangle$  and  $\langle \text{DISTANCE} \rangle$  are first obtained at each zenith angle by clubbing together showers of a particular energy in various core distance bins. Apart from reducing the training data base, following this method also makes ANN training simpler for achieving the desired level of convergence in a reasonable amount of time. The activation function chosen for the present study is the sigmoid function. In order to optimize the number of nodes required in the hidden layer, we also varied the number of the nodes in the hidden layer from 5 to 60 in steps of 5. A plot of the normalised rms error as a function of number of nodes in the hidden layer is shown in Fig 6.6a.

Since increasing the number of nodes beyond 30 results in only a marginal reduction in the normalised rms error (at the cost of longer computation time), it seems one hidden layer with 30 nodes is quite optimum. The normalised rms error at the end of the training, for 30 nodes in the hidden layer, reaches a value  $\sim 3 \times 10^{-3}$  and variation of the same as a function of number of iterations is shown in Fig 6.6b. It is worth mentioning here that a normalised rms error of  $\sim 2.7 \times 10^{-2}$  was achieved in our previous work [6.19] and the improvement seen in the present work is as a result of using more data during ANN training. In order to ensure that the network has not become ‘over-trained’ [6.30], the ANN training is stopped when the normalised rms error stops decreasing any further (somewhere around 8000 iterations).

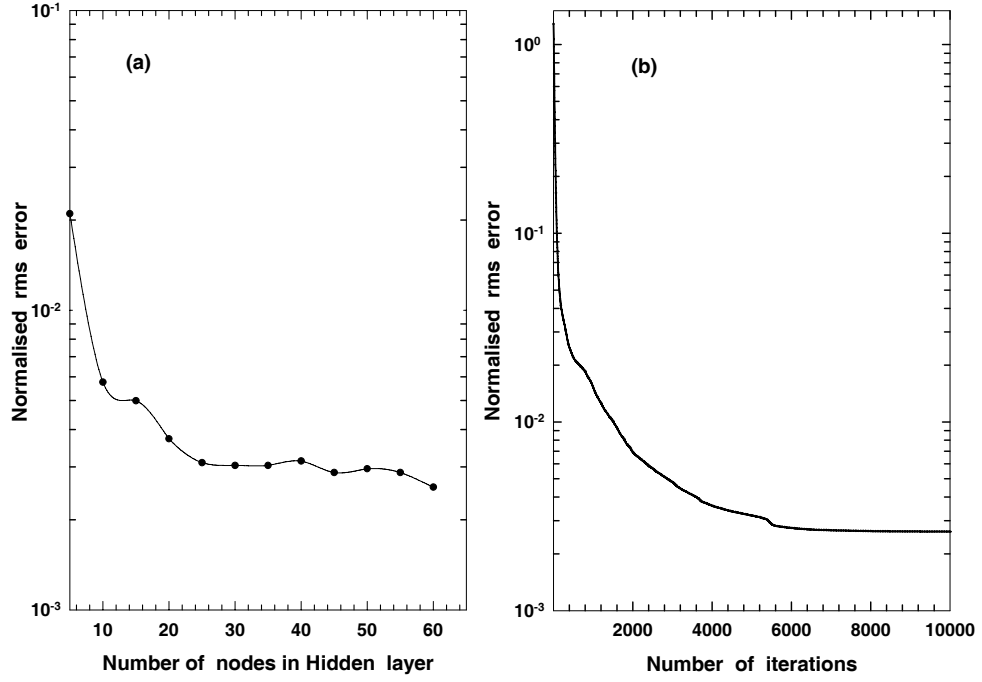


Figure 6.6: (a) Normalised root mean square error as a function of number of nodes in the hidden layer. (b) Normalised root mean square error as a function of number of iterations for 30 nodes in the hidden layer.

### 6.4.2 Testing and validation of the ANN

The ANN is tested with two data samples. The first data sample comprises 10,000  $\gamma$ -ray images (which was earlier used for calculating mean SIZE and mean DISTANCE while preparing the training data set). The second data sample comprises 24,000  $\gamma$ -ray images which were not used at all while preparing the training data set. Both these data samples yielded similar energy reconstruction error plots, thus indicating that ANN has "learned" and not "remembered" the energy reconstruction procedure through over-training [6.30]. Plot of energy reconstruction error obtained for second test data sample is shown in Fig 6.7a. The frequency distribution of  $\Delta_E$  is shown in Fig 6.7b.

It is evident from Fig 6.7a that, the reconstructed energy, employing the ANN method, has a negligible bias in the energy range 1.8 TeV to 20.0 TeV with  $|\Delta_E| < 5.0\%$ . The rms width of the frequency distribution is found to be  $\sim 26\%$  and the same for best fit Gaussian distribution is  $\sim 22\%$ .

The interpolation capability of the ANN-based energy reconstruction procedure, at intermediate  $\gamma$ -ray energies and zenith angles, has also been checked by applying

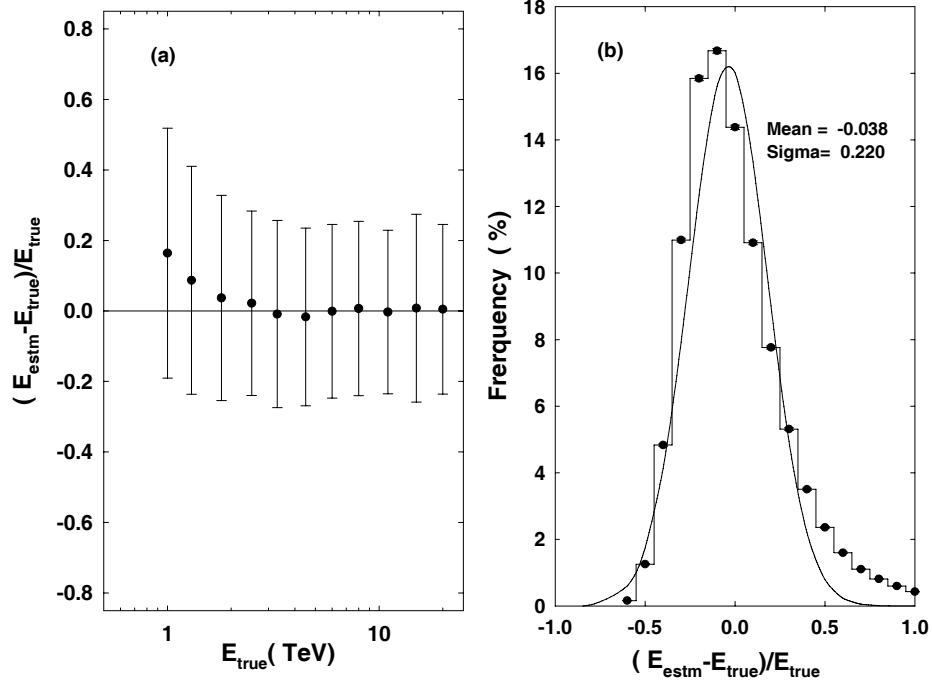


Figure 6.7: (a) Relative mean error in the reconstructed energy ( $\Delta_E = (E_{\text{estm}} - E_{\text{true}})/E_{\text{true}}$ ) as a function of energy for the ANN-based energy estimation procedure. (b) Frequency distribution of  $\Delta_E$  along with a best fit Gaussian distribution to the histogram.

it to an independent validation data base of 4000 showers. The energy of the primary  $\gamma$ -rays was chosen to be 1.1 TeV, 2.1 TeV, 5.2 TeV and 9.5 TeV at zenith angles of  $10^\circ$  and  $20^\circ$ , and 2.1 TeV, 5.2 TeV, 9.5 TeV and 17.0 TeV at zenith angles of  $30^\circ$  and  $40^\circ$ . Since no simulated data at these zenith angles and energies was used during training of the ANN, the results obtained now on the validation data obviously indicate the interpolation capability of the ANN. A plot of energy reconstruction error obtained for the validation data sample is shown in Fig 6.8a. The frequency distribution of  $\Delta_E$  along with a best fit Gaussian distribution to the histogram is shown in Fig 6.8b.

The rms width of the best fit Gaussian distribution for the test and validation data ( $\sim 22\%$  and  $\sim 26\%$ , respectively), with a negligible bias in the energy suggests that the performance of the ANN-based method for energy reconstruction is quite reliable. Taking higher of the two  $\sigma(\Delta_E)$  values (i.e  $\sigma(\Delta_E) \sim 26\%$ ) as a safe value of the energy resolution achieved by the ANN-based for energy reconstruction procedure, one can easily conclude that the proposed method, apart from

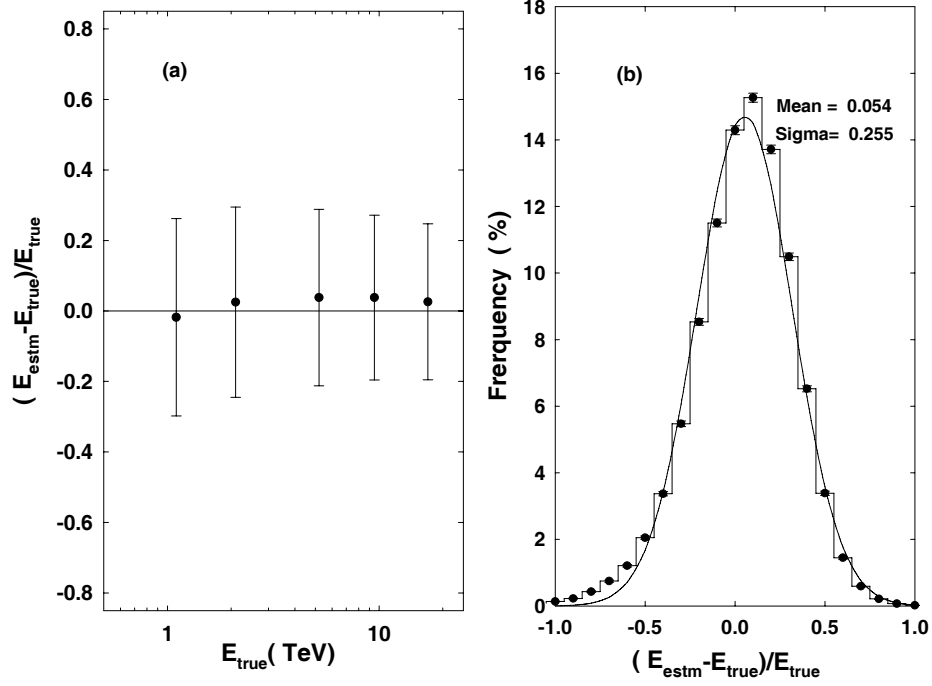


Figure 6.8: (a) Relative mean error in the reconstructed energy ( $\Delta_E = (E_{\text{estm}} - E_{\text{true}})/E_{\text{true}}$ ) as a function of energy for the ANN-based energy estimation procedure when applied to a validation data sample. (b) Frequency distribution of  $\Delta_E$  along with a best fit Gaussian distribution to the histogram.

yielding a comparable performance to that of other single imaging telescopes ( e.g.  $\sigma(\ln E)$  of  $\sim 25\%$  reported by the Whipple group [6.7]), has the added advantage that it considers zenith angle dependence of SIZE and DISTANCE parameters as well. The procedure thus allows data collection over a much wider zenith angle range as against a coverage of upto  $30^\circ$  in case the zenith angle dependence is to be ignored. On analyzing Figures 6.3, 6.4, 6.5 and 6.7, it is also obvious that the ANN-based energy reconstruction procedure yields better results (both in terms of bias and energy resolution) as compared to the conventional energy reconstruction methods. Even though the look-up table based energy reconstruction procedure appears to be equally competitive it suffers from other drawbacks. Implementation of this method requires cumbersome tabulation of interpolated data at a number of DISTANCE, energy and zenith angle values and energy reconstruction procedure is also more time consuming as compared to the ANN method. Implementation of the ANN-based energy reconstruction procedure, on the other hand, is relatively much more straight forward. Once satisfactory training of the ANN is achieved,

the corresponding ANN generated weight-file can be easily used by an appropriate subroutine of the main data analysis program for determining the energy of  $\gamma$ -ray like events. Use of a dedicated ANN software package is thus necessary only during the training of the ANN. Hence, compared to the conventional methods, the ANN-based energy reconstruction procedure offers several advantages like reasonably good energy resolution, applicability over a wider zenith angle range and implementation ease.

## 6.5 Energy spectrum of the Crab Nebula as measured by the TACTIC telescope

In order to test the validity of the ANN-based energy estimation procedure, we have applied this procedure for determining the energy spectrum of the Crab Nebula. For this purpose we reanalyzed the Crab Nebula data collected by the TACTIC imaging telescope for  $\sim 101.44$  h during Nov. 10, 2005 - Jan. 30, 2006. The zenith angle of the observations was  $\leq 45^\circ$ . The data has been collected with inner 225 pixels ( $\sim 4.5^\circ \times 4.5^\circ$ ) of the full imaging camera with the innermost 121 pixels ( $\sim 3.4^\circ \times 3.4^\circ$ ) participating in the trigger. The data recorded by the telescope was corrected for inter-pixel gain variation and then subjected to the standard two-level image 'cleaning' procedure with picture and boundary thresholds of  $6.5\sigma$  and  $3.0\sigma$ , respectively. The clean Cherenkov images were characterized by calculating their standard image parameters like LENGTH, WIDTH, DISTANCE, ALPHA, SIZE and FRAC2 [6.23-6.24]. Before determining the energy spectrum, the agreement between the predictions from Monte Carlo simulations and the actual performance of the telescope has been checked. This is done by comparing the observed trigger rate of the telescope with the predicted value and by comparing the expected and observed image parameter distributions for protons [6.31]. Reasonably good matching is seen between the experimentally observed quantities and those predicted by simulations.

The standard Dynamic Supercuts procedure [6.5,6.19] is used to separate  $\gamma$ -ray like images from the background cosmic-rays. The Dynamic Supercuts  $\gamma$ -ray selection criteria used in the present analysis are slightly less tight than the ones used by us in our earlier work [6.23,6.24] as the main aim here is to increase the number of  $\gamma$ -ray like events with only a marginal loss of statistical significance. The new cut values used for the present analysis are the following :  $0.11^\circ \leq LENGTH \leq (0.260 + 0.0265 \times \ln S)^\circ$ ,  $0.06^\circ \leq WIDTH \leq (0.110 + 0.0120 \times \ln S)^\circ$ ,  $0.52^\circ \leq DISTANCE \leq 1.27^\circ \cos^{0.88} z$ ,  $SIZE \geq 450 d.c$  ( where 6.5 digital counts  $\equiv 1.0$  pe ),  $ALPHA \leq 18^\circ$  and  $FRAC2 \geq 0.35$ . The number of  $\gamma$ -ray events obtained after applying the above cuts are determined to be  $\sim (928 \pm 100)$  with a statistical significance of  $\sim 9.40\sigma$ . Defining  $ALPHA \leq 18^\circ$  as the  $\gamma$ -ray domain and  $27^\circ \leq ALPHA \leq 81^\circ$  as the background region, the number of  $\gamma$ -ray events have been

calculated by subtracting the expected number of background events (calculated on the basis of background region) from the  $\gamma$ -ray domain events.

The differential photon flux per energy bin has been computed using the formula

$$\frac{d\Phi}{dE}(E_i) = \frac{\Delta N_i}{\Delta E_i \sum_{j=1}^5 A_{i,j} \eta_{i,j} T_j} \quad (6.4)$$

where  $\Delta N_i$  and  $d\Phi(E_i)/dE$  are the number of events and the differential flux at energy  $E_i$ , measured in the  $i$ th energy bin  $\Delta E_i$  and over the zenith angle range of  $0^\circ$ - $45^\circ$ , respectively.  $T_j$  is the observation time in the  $j$ th zenith angle bin with corresponding energy-dependent effective area ( $A_{i,j}$ ) and  $\gamma$ -ray acceptance ( $\eta_{i,j}$ ). The 5 zenith angle bins ( $j=1-5$ ) used are  $0^\circ$ - $10^\circ$ ,  $10^\circ$ - $20^\circ$ ,  $20^\circ$ - $30^\circ$ ,  $30^\circ$ - $40^\circ$  and  $40^\circ$ - $50^\circ$  with effective collection area and  $\gamma$ -ray acceptance values available at  $5^\circ$ ,  $15^\circ$ ,  $25^\circ$ ,  $35^\circ$  and  $45^\circ$ . The number of  $\gamma$ -ray events ( $\Delta N_i$ ) in a particular energy bin is calculated by subtracting the expected number of background events, from the  $\gamma$ -ray domain events. The  $\gamma$ -ray differential spectrum obtained after using appropriate values of effective collection area and  $\gamma$ -ray acceptance efficiency (along with their energy and zenith angle dependence) is shown in Fig 6.9a.

A power law fit ( $d\Phi/dE = f_0 E^{-\Gamma}$ ) to the measured differential flux data with  $f_0 \sim (3.12 \pm 0.48) \times 10^{-11} \text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$  and  $\Gamma \sim 2.69 \pm 0.14$  is also shown in Fig 6.9a. The fit has a  $\chi^2/dof \sim 3.64/6$  with a corresponding probability of  $\sim 0.72$ . The errors in the flux constant and the spectral index are standard errors. Reasonably good matching of this spectrum with that obtained by the Whipple and HEGRA groups [6.32-6.33] reassures that the procedure followed by us for obtaining the energy spectrum of a  $\gamma$ -ray source is quite reliable. The confidence ellipses in the two parameters jointly (i.e  $f_0$  and  $\Gamma$ ) at 68.3%, 90%, 95.4% and 99% confidence levels are shown in Fig 6.9b. The corresponding  $\Delta\chi^2$  values of these 4 contours, for 6 degrees of freedom are  $\sim 7.04$ ,  $\sim 10.6$ ,  $\sim 12.8$  and  $\sim 16.8$ .

It is important to point out here that for background cosmic-ray events, which are not coming from the source direction and are classified as  $\gamma$ -ray like events by the Dynamic Supercuts procedure, a wrong energy value will be obtained. However, subtraction of the background events (estimated from  $27^\circ \leq \text{ALPHA} \leq 81^\circ$  region), from the  $\gamma$ -ray domain (defined as  $\alpha \leq 18^\circ$ ), will cancel out these events (in a statistical sense). Estimating the energy spectrum of  $\gamma$ -rays in the presence of background events by following this approach is well known [6.30] and has been used quite extensively by other groups.

## 6.6 Summary

A novel ANN-based energy estimation procedure, for determining the energy spectrum of a candidate  $\gamma$ -ray source has been presented in this chapter. The procedure



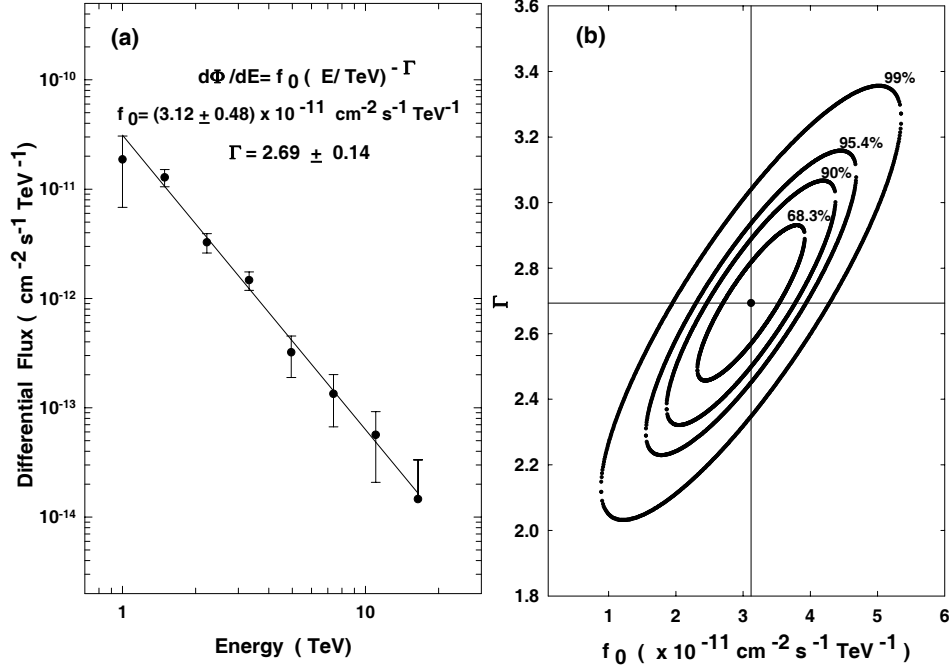


Figure 6.9: (a) The differential energy spectrum of the Crab Nebula as measured by the TACTIC telescope and employing the ANN-based energy reconstruction procedure.(b) The confidence ellipses in the two parameters jointly (i.e  $f_0$  and  $\Gamma$ ) at 68.3%, 90%, 95.4% and 99% confidence levels.

followed by us uses an Artificial Neural Network to estimate the energy of a  $\gamma$ -ray like event on the basis of its image SIZE, DISTANCE and zenith angle. Apart from yielding a reasonably good  $\sigma(\Delta_E)$  of  $\sim 26\%$ , this procedure has the added advantage that it allows data collection over a much wider zenith angle range as against a coverage of upto  $30^\circ$  only in case the zenith angle dependence is to be ignored. We have also successfully implemented the ANN-based energy reconstruction algorithm in our analysis chain, by directly using the ANN generated weight-file, so that the energy of a  $\gamma$ -ray like event could be predicted without using the ANN software package. Reasonably good matching of the Crab Nebula spectrum as measured by the TACTIC telescope with that obtained by the Whipple and HEGRA groups reassures that the procedure followed by us for obtaining the energy spectrum of a  $\gamma$ -ray source is quite reliable.

## 6.7 References

- [6.1] T.C. Weekes, Space Sci. Rev., 75 (1996) 1.
- [6.2] M.F. Cawley et al., Exp. Astron., 6 (1996) 7.
- [6.3] F.A. Aharonian et al., Nature. 440 (2006) 1018.
- [6.4] J. Albert et al., Science. 320 (2008) 1752.
- [6.5] G. Mohanty et al., Astropart. Phys. 9 (1998) 15.
- [6.6] F. Aharonian et al., Astron. Astrophys. 349 (1999) 29.
- [6.7] P.F. Rebillot et al., Ap.J. 641 (2006) 740.
- [6.8] F.A. Aharonian et al., Astropart. Phys. 6 (1997) 343.
- [6.9] F.A. Aharonian et al., Astropart. Phys. 6 (1997) 369.
- [6.10] R. Borisjuk et al., Nucl. Instr. and Meth. A. 381 (1996) 512.
- [6.11] J. Damgov and L.Litov, Nucl. Instr. and Meth. A. 482 (2002) 776.
- [6.12] P.V.M. da Silva and J.M. de Seixas, Nucl. Instr. and Meth. A. 559 (2006) 124.
- [6.13] R.A. Vazquez, F.Halzen and E.Zas, Phy. Rev. D. 45 (1992) 356.
- [6.14] P.T. Reynolds, Irish Astron. J. 21 (1993) 118.
- [6.15] P.T. Reynolds and D.J.Fegan, Astropart. Phys. 3 (1995)137.
- [6.16] P. Boinee, A.De Angelis and G.L.Foresti, Proc. World Academy of Science, Engineering and Technology, 7 (2005) 394.
- [6.17] D. Dumora et al., Nucl. Phys. B ( Proc. Suppl.) 97 (2001) 255.
- [6.18] R.K. Bock et al., Nucl. Instr. and Meth. A. 516 (2004) 511.
- [6.19] K.K. Yadav et al., Astropart. Phys. 27 (2007) 447.
- [6.20] S.V. Godambe et al., J.Phys. G. Nucl. Part. Phys. 35 (2008) 065202.
- [6.21] D. Heck et al., Report FZKA 6019 Forshungszentrum, Karlshrue. (1998).
- [6.22] A.M. Hillas, Proc. 19th. ICRC, La Jolla. 3 (1985) 445.
- [6.23] D.J. Fegan, Space Science Reviews. 75 (1996) 137.

- [6.24] M. Lemoine-Goumard, B.Degrang and M.Tluczykont, *Astropart. Phys.* 25 (2006) 195.
- [6.25] F. Aharonian et al., *Astron. Astrophys.* 457 (2006) 899.
- [6.26] W. Hofmann et al., *Astropart. Phys.* 12 (2000) 207.
- [6.27] T. Schweizer, Ph.D Thesis, Universitat Autònoma de Barcelona, (2002).
- [6.28] J.M. Zurada, *Introduction to Artificial Neural Systems*, Jaico Publishing House, Mumbai, 2006.
- [6.29] M. Reidmiller, *Computer Standards Interfaces.* 16 (1994) 265.
- [6.30] R.O. Duda, P.E.Hart and D.G.Stork, *Pattern Classification*, John Wiley And Sons, New York, 2002.
- [6.31] R. Koul et al., *Nucl. Instr. and Meth. A.* 578 (2007) 548.
- [6.32] A.M. Hillas et al., *Ap.J*, 503 (1998) 744.
- [6.33] F.A. Aharonian et al., *Ap.J*, 614 (2004) 897.

# Chapter 7

## Conclusions and future work

### 7.1 Brief summary of the work presented

The field of VHE  $\gamma$ -ray astronomy has produced many exciting results over the last decade largely due to the development of the ground based Cherenkov imaging techniques pioneered by the Whipple collaboration. This technique was applied for the first time in 1988-1989 by the Whipple collaboration to detect steady  $\gamma$ -ray emission from the Crab Nebula using a 10m reflector with a 37- pixel photomultiplier tube camera [7.1]. Another important development in the field is the application of stereoscopic imaging, proposed by the HEGRA group [7.2], which proves that the information available from multiple telescopes, located within the same Cherenkov light pool, can dramatically improve the sensitivity of the technique. The instruments use multiple telescopes to image the air-shower from different viewing angles for improved reconstruction of  $\gamma$ -ray direction and rejection of cosmic-ray background. The stereoscopic method also allows rejection of triggers caused by cosmic-ray secondary muons. It is thus imperative that telescope systems such as H.E.S.S., for example, provide an angular resolution for single  $\gamma$ -rays of 3 arc-min to 6 arc-min, a  $\gamma$ -ray energy resolution of around 15% and an unprecedented sensitivity with cosmic-ray rejection factor of more than 99.9%. It is the landmark development of these two methods that revolutionized the field of ground-based very high-energy (VHE)  $\gamma$ -ray astronomy. In addition to the ground based methods, the success of the satellite borne experiments like CGRO and the more recent experiments like the Fermi-LAT have revealed the rich nature of the  $\gamma$ -ray sky and provided a guide for TeV observations. Apart from the discovery of the Crab Nebula as a standard candle, the discovery of VHE emission from the active galactic nuclei like the Markarian 421, Markarian 501, PKS 2155-304, PKS 1222+21, PKS 1510-089 etc. is particularly exciting, due to the observed rapid variability of these sources at TeV energies.

The recent exciting results include the ones from the galactic plane survey. The region between  $\pm 30^\circ$  in longitude and  $\pm 3^\circ$  in latitude relative to galactic center has

been extensively observed at an average flux sensitivity of  $\sim 2\%$  of the Crab Nebula for point source searches at energies above 200 GeV. The inner part of our galaxy now contains several known sources of VHE  $\gamma$ -rays. The increased number of sources will allow the first ever study of the behaviour of populations of sources in this energy band. A major task of multi-wavelength follow-up observations now lie ahead to understand the processes at work in these exotic astrophysical environments. The paradigm of cosmic-ray acceleration in SNRs is consistent with these new findings. Follow-up observations of the new sources are planned with different telescopes as well as an extension of this survey into unexplored regions of the galactic plane [7.3]. Another interesting aspect of these sources is that nearly one-third are still not clearly identified with objects seen at other wavelengths, thus having a precise angular resolution is another challenge. Coupled with this, a lower flux threshold is needed to provide a greater sample of SNRs, in particular those which are not biased by being just above the detection threshold. The spectrum will also have to be taken to an energy threshold of well below 50 GeV to overlap with the satellite regime. The future multi-wavelength studies of all these objects promise to provide invaluable information about the particle acceleration mechanisms involved in these source environments. The imaging atmospheric Cherenkov technique has thus opened up a new but challenging window of the electromagnetic spectrum which is now beginning to provide insights into very high energy astrophysical environments and the origin of cosmic-rays. Ground-based imaging atmospheric Cherenkov telescopes have become the most efficient instruments for the observation of  $\gamma$ -rays in the TeV energy range [7.4-7.6]. These  $\gamma$ -rays trace back to cosmic accelerators which can be responsible for the creation and acceleration of the charged cosmic-rays, and are expected to come from a wide variety of cosmic objects within and outside our galaxy. Studying this radiation in detail can yield valuable and unique information about the unusual astrophysical environment characterizing these sources, as also on the intervening intergalactic space [7.7-7.9].

In the beginning of the thesis, a general overview of ground based  $\gamma$ -ray astronomy, in general and atmospheric Cherenkov technique in particular has been presented. The progress and milestones achieved in the field, difficulties of the analysis methods, challenges in the field and the current observational status of the field has also been discussed in detail. Since the thesis revolves around the optimization of the new ANN algorithms and their subsequent use for improving the sensitivity of the TACTIC telescope, it thus becomes critical to provide a detailed introduction about the telescope. The details of the TACTIC telescope, its mechanical structure, optical geometry, electronics used and summary of the performance evaluation of the telescope has also been presented. The telescope has been in operation at Mt. Abu, India and has so far detected  $\gamma$ -ray emissions from the Crab Nebula, Mrk 421 and Mrk 501 and we have been able to give upper limits on 1ES2344 + 514, IC310, H1426 + 428 etc. While excellent matching of the Crab Nebula spectrum with that obtained by other groups reassures us that the

telescope subsystems are functioning properly, there is however a need to improve the present sensitivity level which stands at the detection of 1 Crab Unit at  $5\sigma$  in 25h. Also, apart from validating the stability of the TACTIC subsystems directly with  $\gamma$ -rays from this source, matching of the Crab Nebula spectrum also validates the full analysis chain, including the inputs used from the Monte Carlo simulations, like the effective area and  $\gamma$ -ray acceptance factors and the energy reconstruction procedure. Furthermore, keeping in view that the weak signature of possible cutoff in the energy spectrum of Mrk 421, inferred from our observations which is fairly consistent with the observations of other groups [7.10-7.11], we believe that there is considerable scope for the TACTIC telescope to monitor similar TeV  $\gamma$ -ray emission activity from other active galactic nuclei on a long term basis.

Since the aim of the present thesis is essentially to show the power and advantages of applying an intelligent machine learning algorithm like ANN, it is also imperative to have an indepth appreciation related to the field of ANN, its working methodology, various algorithms contained in the ANN umbrella, the supervised, the unsupervised and the recurrent methods along with the algorithms contained therein with special emphasis on the newer techniques used by researchers and also the merits and demerits of the ANN usage. Having discussed so, it becomes crucial to have an intercomparison of these varied methods to have a feel of the strengths/limitations of every single algorithm. While comparative performance of some ANN algorithms like standard backpropagation, fuzzy logic, genetic algorithms, fractals etc., has been studied for various applications, a rigorous intercomparison of some of the powerful algorithms (e.g. the ones studied in this thesis) was missing from the literature. The work thus also provides a rigorous comparative study of various powerful algorithms, by first applying them to standard benchmark problems and then applying as a regression tool for approximating functions like  $\cos(x)$  and a few special functions. Results of the study conducted suggested that while Levenberg-Marquardt ANN algorithm yields the lowest RMS error for the N-bit Parity and the Two Spiral problems, Higher Order Neurons ANN algorithm gives the best results for the IRIS data problem. The best results for the XOR problem are obtained with the Neuro Fuzzy ANN algorithm. Also, an attempt has been made for the first time to apply ANN for solving the regression problems, such as evaluation of special functions like the Gamma function, the complimentary Error function and the upper tail cumulative  $\chi^2$ -distribution function.

Although the comparative performance of different ANN algorithms is in general problem dependent, the study undertaken in the thesis, gives an insight into the power of various powerful ANN algorithms. Since for real world problems, it is not an easy task to identify the most suitable ANN algorithm by just having a look at the problem, the results presented in the thesis suggest that while investigating the comparative performance of other ANN algorithm, the Levenberg-Marquardt ANN algorithm deserves a serious consideration and cannot be rejected outright because of its training time overheads.

After having established the superiority of the Levenberg-Marquardt ANN algorithm compared to the other methods available within the ANN domain, the study has thus been extended to the core problem, the  $\gamma$ /hadron segregation potential of various ANN algorithms for the TACTIC telescope, by applying the ANN based methods to the Monte Carlo simulated and the observation data on the Crab Nebula. The results of the study indicate that the performance of Levenberg-Marquardt based ANN algorithm is somewhat superior to the Dynamic Supercuts procedure especially beyond  $\gamma$ -ray energies of  $\geq 9$  TeV. The main advantage of using the ANN methodology for  $\gamma$ /hadron segregation work is that it is more efficient in retaining higher energy  $\gamma$ -ray events and this has allowed us to extend the TACTIC observed energy spectrum of the Crab Nebula up to an energy of  $\sim 24$  TeV. Reasonably good matching of the Crab Nebula spectrum as measured by the TACTIC telescope with that obtained by the other groups reassures that the ANN-based  $\gamma$ /hadron segregation method is quite reliable. Another advantage of this approach is that once a satisfactory performance with respect to the segregation potential of the ANN is achieved, a ANN weight-file can be generated which can be easily used by an appropriate subroutine of the main data analysis program for determining the  $\gamma$ -ray like events. Use of a dedicated ANN based software is thus necessary only during the training of the ANN and becomes redundant once the training procedure is completed.

Also, having established the advantages of the ANN based method for selecting higher number of  $\gamma$ -rays, a novel ANN-based energy estimation procedure, for determining the energy spectrum of a candidate  $\gamma$ -ray source has also been developed. Since the intensity of the Cherenkov light (i.e image SIZE) is a function of core distance, which is not possible to obtain with a single imaging telescope, the angular distance of the image centroid from the camera center (known as the DISTANCE parameter) is generally used as an approximate measure of the impact distance. The energy reconstruction procedure with a single imaging telescope thus involves using SIZE and DISTANCE parameters of the Cherenkov event for determining energy of the primary  $\gamma$ -ray. Although the method has been found to work reasonably well over a restricted zenith angle range of  $\leq 30^\circ$ , there is a need to include zenith angle dependence in the energy reconstruction procedure for allowing data collection over a much wider zenith angle range. The procedure followed uses an Artificial Neural Network to estimate the energy of a  $\gamma$ -ray like event on the basis of its image SIZE, DISTANCE and zenith angle. This new energy estimation procedure employed, has yielded a reasonably good  $\sigma(\Delta_E)$  of  $\sim 26\%$ . ANN-based energy reconstruction algorithm has successfully been implemented in our data analysis chain, by directly using the ANN generated weight-file, so that the energy of a  $\gamma$ -ray like event could be predicted without using the ANN software package. Reasonably good matching of the Crab Nebula spectrum as measured by the TACTIC telescope with that obtained by the Whipple and HEGRA groups reassures that the procedure followed for obtaining the energy spectrum of a  $\gamma$ -ray source is quite reliable.

## 7.2 Multivariate background rejection and the future of ANN

There are inherently different approaches adopted by space-based and ground based astronomy telescopes for cosmic-ray background rejection. For the ground based telescopes like TACTIC, additional analysis needs to be conducted in order to separate the signal events from the background events. As discussed in the thesis, the existing ground-based Cherenkov systems do this using the standard Hillas method of parameterising the shower images. The method itself is extremely robust, but for TACTIC, in order to enhance its sensitivity additional analysis techniques need to be applied. Application of the Artificial Neural Networks to improve the gamma/hadron segregation methodology, is one such attempt in this direction. By implementing a multivariate method, the information contained in many parameters, that help to distinguish between signal and background events, can be combined to obtain a single response parameter. This response parameter called eta ( $\eta$ ) in the current work enables events to be classified according to their  $\gamma$ -ray or cosmic-ray likeness. When utilising a multivariate methodology we have demonstrated that significant improvements in performance can be obtained compared to the standard background rejection method. Multivariate methods can be seen as an extension of the standard cuts-based discrimination method. The cost effectiveness, availability and processing power of modern desktop computers enables machine learning algorithms such as Artificial Neural Networks, Decision Trees, Likelihood Estimators, Fisher Discriminants etc to be easily implemented. As noted by Ohm et al,[7.12], one of the main advantages of Neural Networks over Likelihood, Fisher discriminants and other such methods, is their ability to handle nonlinear correlations between parameters which has also been amply demonstrated in this thesis. It is also concluded in the thesis that Neural Networks are the preferred multivariate method for two reasons: they ignore non-discriminating parameters and their decision trail is transparent. Neural Networks are therefore implemented in the work presented and they have been proven to provide results with consistent performances. Attempts in the field are now underway to use real-time Neural Networks, though being applied in other fields, the idea is still at infancy in the field of VHE  $\gamma$ -ray astronomy.

In the last few years there has been an increased interest towards the astronomical applications of NNs even though, in spite of the great variety of problems addressed, most applications still make use of handful of neural models only. This situation is bound to change due to the huge increase in both the quality and the volume of data which is becoming available to the astronomical community worldwide. Conservative predictions lead to the expectations that in about 3 – 5 years, nearly 100 TB of data will be acquired worldwide every night and most of this data will become available to the astronomical community. These huge and heterogeneous data sets will open possibilities which so far are just unthinkable, but it is



already clear that their scientific exploitation will require the implementation of automatic tools capable to perform a large fraction of the routine data reduction, data mining and data analysis work.

Historically, the first attempts to use Neural Networks on astronomical data were meant to separate stars from galaxies or to classify galaxy morphologies or spectral features [7.13]. While the problem of star/galaxy classification seems to have been satisfactorily answered, the problems like object detection, due to their intrinsic complexity, are still a matter of research and will be addressed by the astronomical community in the near future. Neural Networks have also been applied to planetary studies [7.14]; to the study and prediction of solar activity and phenomena [7.15], to the study of the interplanetary magnetic field [7.16], and to stellar astrophysics [7.17]. Other fields of application also include: time series analysis [7.18], and the identification and characterization of peculiar objects such as QSO's, ultraluminous IR galaxies, and  $\gamma$ -ray Bursts [7.19], the noise removal in pixel lensing data, the decomposition of multifrequency data simulated for the Planck mission, the search for galaxy clusters etc.

Still in its infancy, is the use of NN for the analysis of the data collected by the new generation of instruments for astroparticle physics such as, for instance, the Solar energetic proton events, the cosmic-ray telescopes AUGER [7.20] and ARGO [7.21]; the  $\gamma$ -ray Cherenkov telescope [7.22], the VIRGO gravitational waves interferometer [7.23] and even for the search of the Higgs boson [7.24].

### 7.3 TACTIC Telescope upgradation and the road ahead

The TACTIC  $\gamma$ -ray telescope has been in operation at Mt. Abu, India for over a decade now to study TeV  $\gamma$ -ray emission from celestial sources. During this time, apart from consistently detecting a steady signal from the Crab Nebula above  $\sim 1.2$  TeV energy, at a sensitivity level of  $\sim 5.0\sigma$  in  $\sim 25$  h, the telescope has also detected flaring activity from Mrk 421 and Mrk 501 on several occasions. We have very recently carried out improvements in the various sub-systems of the telescope which has resulted in a substantial improvement in its detection sensitivity (viz.,  $\sim 5\sigma$  in an observation period of  $\sim 12$ h as compared to  $\sim 25$ h earlier).

Major upgrading work, involved replacement of signal and high voltage cables and installation of new Compound Parabolic Concentrators (CPC) with the aim to improve the sensitivity of the telescope. New CPCs with square entry and circular exit aperture were installed on the TACTIC imaging camera in order to increase its photon collection efficiency. The reflection coefficient of the new CPCs was measured to be  $\sim 85\%$  in the wavelength range  $\sim 400\text{nm}$ - $550\text{nm}$ . Apart from removing the dead space in between the PMTs completely, use of the new CPCs has also helped us to improve the gamma/hadron segregation capability of the

telescope. In addition, the trigger criteria was also modified by including more nearest neighbor collinear triplet combinations. A dedicated CCD camera was also installed for conducting detailed point run calibrations and data collected has been successfully used for determining the position of the source in the image plane with an accuracy of better than  $\sim \pm 3$  arc min. The analysis procedure has also been upgraded by using the Asymmetry parameter so that an additional 40 % of the hadronic background can be further removed by identifying the "head" and "tail" feature of Cherenkov images. It is worth mentioning here that the  $\gamma$ -ray images have their head closer to the assumed source position in the imaging camera and thus can be selected preferentially by imposing  $\text{Asymmetry} > 0$  cut. Analysis of the  $\sim 27$ h data collected with the telescope between 23-Dec-2011 and 29-Dec-2011, after completing the hardware and software upgrading work, yields an increase in the prompt coincidence rate from  $\sim 2.33$  Hz (at zenith angle of  $0^\circ$ ) to  $\sim 3.70$  Hz. This translates to reduction in the threshold energy of the telescope for cosmic-rays from  $\sim 1.8$  TeV to  $\sim 1.4$  TeV and from  $\sim 1.2$  TeV to  $\sim 0.8$  TeV for the  $\gamma$ -rays. Detailed simulation studies are still in progress to confirm these estimates. As a result of the upgradation and more refined data analysis, the sensitivity of the TACTIC telescope has improved from  $N_\sigma \sim 0.97 \sqrt{T}$  to  $N_\sigma \sim 1.40 \sqrt{T}$ . The  $\gamma$ -ray rate per hour from the Crab Nebula has also increased from  $\sim 9.13$  to  $\sim 15.00$ . The telescope can now detect the TeV  $\gamma$ -ray emission from the Crab Nebula at  $\sim 5\sigma$  in an observation period of  $\sim 12$ h as compared to  $\sim 25$ h earlier.

The successful applications of the ANN methodology demonstrated in this thesis will again be taken up in the near future, for gamma/hadron segregation and energy estimation procedures and applied to the new data base generated by the upgraded telescope over the next 1 year or so. This should enhance the telescope sensitivity further which will be invaluable for long term monitoring with the TACTIC telescope.

It is worth having a closer look at some of the scientific drivers which motivate observations in the multi-TeV energy range with a dedicated instrument like TACTIC. With ever growing number of  $\gamma$ -ray sources in the sky and rather limited field of view of Cherenkov imaging telescopes most of the TeV sky remains relatively unexplored and only 10 % of the sky has been observed. Searching for  $\gamma$ -ray emission from different types of blazars including high energy peaked BL Lac objects (HBLs), intermediate energy peaked BL Lac objects (IBLs), and flat spectrum radio quasars (FSRQs) is an important research field where we can also contribute to understand the  $\gamma$ -ray emission characteristics in different types of blazars. The most notable observational results from the blazars have been extremely fast large amplitude flux and spectral variability on hour time scales, and a pronounced X-ray, TeV  $\gamma$ -ray flux correlation. Studying the temporal correlation between flux variations at different wavelengths during flares can help in providing constraints on the blazar emission models. While multiwavelength information, in particular at radio and X-ray energies, can provide constraints on hadronic or electronic components but often one

requires model-dependent assumptions to decide the nature of the parent particles.

One of the objectives of the TACTIC telescope will also be to conduct deep observations on a small number of well known VHE blazars like Mrk421, Mrk501, H1426+428, 1ES1959+650, BL Lacertae and 1ES 2344+514. In this regard, it is quite reassuring to find that the energy spectrum of the Mrk421 measured by the TACTIC telescope, for the data collected between December 27, 2005 and Feb 07, 2006, has been used while interpreting the results of the Swift observations between April 2006 and July 2006. Participating in target of opportunity observations of flaring blazars is another important area where results from the TACTIC telescope can help in constraining the blazar emission models. Further, the TeV energy spectra of extragalactic sources also carry the imprint of extragalactic absorption in the form of ‘characteristic high-energy cutoffs’ at energies above 100 GeV energy range. Detecting this imprint of absorption thus is yet another scientific objective where results from TACTIC telescope can be useful by providing deep observations of blazars and by determining their energy spectra with good accuracy.

## 7.4 Other possible applications of ANN

### 7.4.1 Image cleaning with ANN

Apart from this, at the cleaning level itself nearly 30-40%  $\gamma$ -ray images are lost. The traditional cleaning method used to analyse images provided by Cherenkov telescopes selects those pixels which are above the ‘picture’ threshold or are beside such pixels which have signal above the lower ‘boundary’ threshold. A significant quantity of the total number of PMTs that have real signal are rejected by this method. The percentage can be increased by lowering the picture and boundary thresholds. However, this in turn also increases the number of noise pixels incorrectly selected as signal pixels which makes cleaning an extremely difficult task. We also propose to use an ANN based algorithm which will be trained to select the pixels depending on whether they are a part of the genuine signal or not. The method will not involve any image cleaning at all and shall thus be inherently more powerful in accepting more genuine signals.

### 7.4.2 Application of ANN to raw images

Having seen the potential of an ANN based algorithm for segregating the  $\gamma$  /hadron showers, work will be undertaken to extend this study further to evaluate the possibility of using the ANN for segregation of  $\gamma$  /hadron on the basis of raw Cherenkov images. For this purpose, we plan to use raw simulated images as would have been recorded by the 349-pixel camera of the TACTIC telescope, at  $\gamma$ -ray energies of 1.0-27 TeV. Simulated proton energies will be taken between the energy range of 2.0 -40 TeV. A total base of  $\sim 50,000$   $\gamma$ -rays and an equal set of proton images will

be generated, from which  $\sim 33\%$  of the data will be used for training the ANN and the remaining  $\sim 67\%$  of the data will be used for testing of the network. The ANN configuration will have 349 inputs corresponding to each pixel output and the output will be 1 node depending on whether the event in question is  $\gamma$ -ray like or a proton event. To reduce the complexity of the problem, we may also use some data compressing technique like PCA etc, so that the number of inputs can be reduced without compromising on the information contained in the inputs.

## 7.5 New Experiments

### 7.5.1 MACE Telescope

The current generation of imaging atmospheric Cherenkov telescopes (IACT) has increased the number of known GeV/TeV astrophysical sources to more than 160. The Himalayan Gamma Ray Observatory (HiGRO) is the highest altitude ground based  $\gamma$ -ray observatory using the atmospheric Cherenkov technique, located at Hanle ( $32.8^\circ$  N;  $78.9^\circ$  E; 4200 m asl) in the Ladakh region of Himalayas, in Northern India. High Altitude GAMMA Ray (HAGAR) telescope which is a wavefront sampling array of seven telescopes each with 7 para-axially mounted mirrors of 0.9m diameter, has been in operation there since 2008 as the first phase of the HiGRO. The MACE (Major Atmospheric Cherenkov Experiment) telescope [7.25] which is currently being setup there will explore the  $\gamma$ -ray sky at energies down to 30GeV. The 155 ton MACE telescope follows a track and wheel design and is supported on 6 wheels of 60 cm diameter which move on a 27m diameter track. The 21m diameter light collector of the telescope is made up of honeycomb based spherical Aluminium mirror facets (size  $50\text{cm} \times 50\text{cm}$ ) which have been manufactured within the country. They have graded focal length between 25.1m and 26.2m and produce an on-axis spot size of  $\sim 15\text{mm}$  diameter. The imaging camera of the telescope comprises 1088 pixels covering a total field-of-view of  $\sim 4.3^\circ \times 4.0^\circ$ . The signal processing and data acquisition electronics is housed in the camera shell itself. The imaging camera follows a modular design with 16 pixels forming a Camera Integrated Module (CIM) complete with its signal processing and digitization electronics. The data from all the 68 modules are integrated in a data concentrator before being sent to archiving and display systems located in the control room. The structural elements of the telescope are presently being assembled at the manufacturer's facility in Hyderabad. The drive system tests are likely to be started soon after installing  $\sim 30$  mirror panels (1 panel has 4 mirrors to form a  $1\text{m}^2$  area) on the space frame of the telescope. The modular data acquisition hardware and software is at an advanced stage of prototyping. After detailed testing the telescope is planned to be shifted to Hanle by mid 2014 and we expect the first engineering runs to commence at Hanle by early 2015.



Figure 7.1: Present status of MACE telescope.

### 7.5.2 Cherenkov Telescope Array (CTA)

The CTA project [7.26-7.27] is an initiative to build the next generation ground-based very high energy  $\gamma$ -ray instrument. It will serve as an open observatory to a wide astrophysics community and will provide a deep insight into the non-thermal high-energy universe. The proposed CTA is a large array of Cherenkov telescopes of different sizes, based on proven technology and deployed on an unprecedented scale. It will allow significant extension of our current knowledge in high-energy astrophysics. CTA is a new facility, with capabilities well beyond those of conceivable upgrades of existing instruments such as H.E.S.S., MAGIC or VERITAS. CTA will, for the first time in this field, provide open access via targeted observation proposals and generate large amounts of public data, accessible using Virtual Observatory tools. CTA aims to provide full-sky view, from a southern and a northern site, with unprecedented sensitivity, spectral coverage, angular and timing resolution, combined with a high degree of flexibility of operation. CTA will be about a factor of 10 more sensitive than any existing instrument. It will therefore for the first time allow detection and in-depth study of large samples of known source types, will explore a wide range of classes of possible  $\gamma$ -ray emitters beyond the sensitivity of current instruments, and will be sensitive to new phenomena. In its core energy range, from about 100 GeV to several TeV, CTA will have milli-Crab sensitivity, a factor of 1000 below the strength of the strongest steady sources of VHE  $\gamma$ -rays, and a factor of  $\sim 10000$  below the highest fluxes measured in bursts. This dynamic

range will not only allow study of weaker sources and of new source types, it will also reduce the selection bias in the taxonomy of known types of sources.

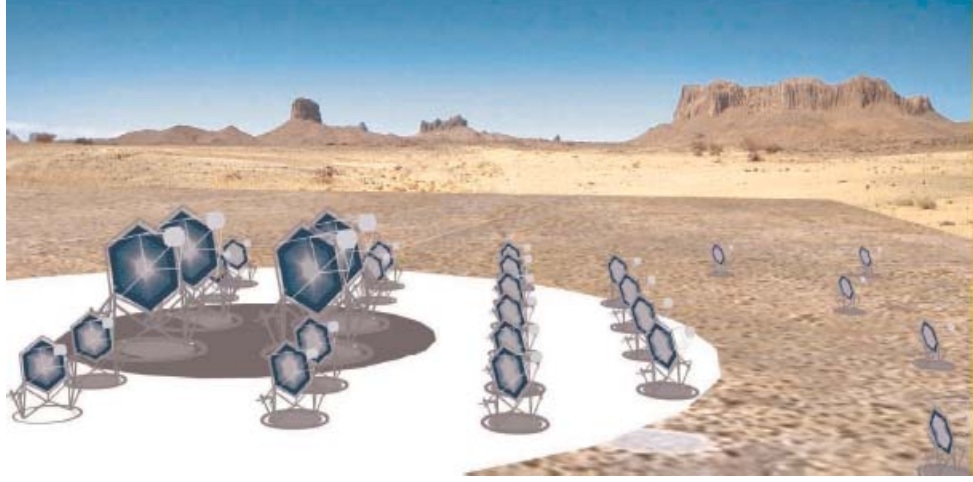


Figure 7.2: Conceptual layout of the Cherenkov telescope array.

## 7.6 New insights to VHE $\gamma$ -ray observations

TeV  $\gamma$ -rays from dark matter condensing around the galactic center is an exciting new prospect for TeV  $\gamma$ -ray astronomy. However it is not easy to obtain exclusively a conclusion that VHE  $\gamma$ -rays from galactic center are due to the dark matter origin and not from conventional emission mechanism. A progress to some extent is expected from newly starting projects by mapping the region of emission with good angular resolution and by watching the time variability, but it is important to improve the IACT's to have, in particular better energy resolution.

Since the Galactic Center is likely to have a large irreducible  $\gamma$ -ray background, if the sensitivity of future instruments can be improved to a level of  $\sim 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$  then a number of different sources ranging from very nearby galactic substructure (microhalos) to the closest galactic clusters can be detected. Close to the Earth itself, there is the possibility that the galactic halo has a significant substructure that

might be observable. Outside of our Galaxy, nearby low surface brightness galaxies and dwarf spheroidals such as Ursa Minor, Draco and Willman1 are possible observables.

The extragalactic background light (EBL) consists of photons emitted by galaxies over the history of the universe. Interactions between EBL photons and  $\gamma$ -rays can obscure  $\gamma$ -ray sources, and produce a link between the history of galaxy formation and VHE observations. Direct measurement of the EBL is difficult, because the EBL is produced across cosmological time from all types of galaxies, sophisticated modeling is required to understand the buildup of this photon population. Modeling the sources of the EBL has been done historically using a variety of techniques, including evolution of galaxy properties that are either inferred over some range in wavelength or directly observed in galaxy surveys. Photon-photon interactions with sufficient energy can create electron-positron pairs. For multi-GeV and TeV  $\gamma$ -rays, interactions with UV and IR background photons can create an optically-thick barrier to passage over cosmological distances. This effect has been used by many authors in recent years to constrain the EBL and disfavor or exclude specific models.

In addition to the above, future generation  $\gamma$ -ray telescopes will mainly be driven to better address the particle accelerations, to address the origin of cosmic radiation, to detect the annihilation and production of dark matter, to understand the  $\gamma$ -ray bursts and finally to enable the discovery of new types of sources unseen at other wavelengths. Having said so, the future of VHE  $\gamma$ -ray astronomy therefore is extremely promising as the goals to be addressed are limitless.

## 7.7 References

- [7.1] T.C. Weekes et al., Ap.J 342 (1989) 379.
- [7.2] A. Daum et al., Astropart. Phys. 8 (1997) 1.
- [7.3] A.M. Hillas, Astropart. Phys 43 (2013) 19.
- [7.4] J. Holder, Astropart. Phys. 39 (2012) 61.
- [7.5] F. Aharonian et al., Rep. Prog. Phys. 71 (2008) 096901.
- [7.6] P.M. Chadwick et al., J.Phys. G : Nucl. Part. Phys. 35 (2008) 033201 .
- [7.7] F.W. Stecker and O.C.De Jager, Ap.J 476 (1997) 712.
- [7.8] E. Dwek and F. Krennrich, Ap.J 618 (2005) 657.

- [7.9] F.Aharonian et al., Nature 440 (1989) 1018.
- [7.10] F.Krennrich et al., Ap.J, 560 (2001) L45.
- [7.11] F.Aharonian et al., A&A., 437 (2005) 95.
- [7.12] S. Ohm et al., Astropart. Phys. 31 (2009) 383.
- [7.13] R.K. Gulati et al., Artificial Neural Networks in Stellar Astronomy (2001).
- [7.14] K. Lepper et al., Lumin. Measur. and Neural Network Analysis, 33-rd Ann. Lunar and Planet. Sci. Conf., Abs. 1400] (2002).
- [7.15] F. Borda et al., 347 Solar Physics, (2002) 206.
- [7.16] I. S. Veselovskii et al., Astronomicheskii Vestnik, 116 (2000) 34.
- [7.17] O. Fuentes, Experimental Astronomy, 21 (2001) 12.
- [7.18] R. Tagliaferri et al., Heidelberg: Springer-Verlag, 386 (2001).
- [7.19] A. Balastegui et al., arXiv:0108272v1 (2001).
- [7.20] G.A.M Tanco et al., arXiv:0109140v1 (2001).
- [7.21] M. Amenomori et al., arXiv:0010292v1 (2000).
- [7.22] B.M. Schaefer et al., arXiv:0101318v1 (2001).
- [7.23] F. Barone et al., Springer-Verlag, 171 (1999).
- [7.24] D. Smirnov, Proc. of the Meeting of the American Physical Society (2002).
- [7.25] R. Koul et al., Proc. 32rd ICRC, Beijing, China. (2011) OG 2.5-803.
- [7.26] The CTA consortium, Exp Astron. 32 (2011) 193.
- [7.27] B.S. Acharya et al., Astropart. Phys 43 (2013) 3.



# Appendix A

## Gamma-ray production mechanisms

Gamma-rays are produced in non-thermal processes like interactions of radiation with matter fields. Supernova remnants, the regions where charged particles may be accelerated to TeV energies at the shock fronts, can produce such  $\gamma$ -rays. Also the vicinity of a neutron star which is highly magnetized, or jets of an AGN are possible production sites of these high-energy  $\gamma$ -rays. In the following sections, the possible production mechanisms of high energy  $\gamma$ - rays are briefly summarized.

- **Charged particles in strong electric or magnetic fields**

The charge of a particle at rest produces a Coulomb field. When the particle moves, its corresponding electromagnetic field also varies. According to Maxwells equations, all accelerated charged particles emit electromagnetic radiation. photons are emitted by accelerated charged particles, while momentum is conserved in the whole process.

- **Cyclotron process**

When a non-relativistic charged particle enters into a magnetic field, it gyrates (rotates) non-relativistically around magnetic field lines with an angle  $\theta$  (pitch angle) between the particle's trajectory and the direction of magnetic field and with a specific Larmor frequency given by:

$$\nu_L = eB/mc \tag{A.1}$$

where  $e$  and  $m$  are the charge and the mass of the particle, respectively.  $B$  is the magnetic field strength and  $c$  is the speed of light. The gyration radius is maintained by the balance between the Lorentz force of the magnetic field and the centrifugal repulsion of the orbiting particle. A rotating charged particle emits electromagnetic waves. This type of radiation is called cyclotron radiation. It is also observed that while the charged particle is moving in

the magnetic field, circularly or linearly polarized electromagnetic waves are emitted depending on the direction of the observer to the magnetic field.

- **Synchrotron radiation**

Cyclotron radiation is replaced by the synchrotron radiation, when the charged particle moves with a speed close to the speed of light. The motion of the particle is circular in trajectory and uniform around the magnetic field lines, but if the velocity along the field lines is non-zero, the path becomes helical as shown in Fig A.1. Therefore, the radiation emitted by the charged particles is beamed into a cone of angle  $v = mc^2/E$ . An observer located at the orbital plane of the electron will only see radiation when the cone is pointed in that direction. Instead of a single frequency, the radiation now is emitted as a continuum spectrum about the  $\nu_c$ , which is the critical frequency at which the maximum power is emitted.  $\nu_c$  can be written as

$$\nu_c = \frac{3}{2} \left( \frac{eB}{mc} \right) \Gamma^2 \sin\phi \quad (\text{A.2})$$

where  $\phi$  is the pitch angle between the direction of the magnetic field and that of the electron and  $\Gamma = E/m$  is the Lorentz factor of the particle with mass  $m$  and energy  $E$ .

The loss of energy is given by

$$-\frac{dE}{dx} = \frac{1}{c} \frac{dE}{dt} = (2e^4/3m^2c^4) \Gamma^2 B^2 \text{erg cm}^{-1}. \quad (\text{A.3})$$

where  $B$  is measured in gauss and  $E$  in erg.

The power emitted by an accelerated particle has a characteristic two-lobe distribution around the direction of the acceleration. In astrophysical sources the electron energies are obeying a power-law with index  $\alpha$  so that  $N(E) \propto E^{-\alpha}$ , then the synchrotron spectrum also follows a power-law of  $P(\nu) \propto \nu^\beta$ , where the spectral index is  $\beta = (1 - \alpha)/2$ .

- **Curvature Radiation**

In a strong magnetic field ( $\sim 10^{12}$  G) an electron may be constrained to follow the path of a magnetic field line very closely, with pitch angle nearly zero. The magnetic field lines are generally curved and the electrons are accelerated transversely and begin to radiate. This radiation is called curvature radiation. The frequency spectrum of curvature radiation is like the spectrum of synchrotron radiation: the spectrum depends on the magnetic field strength, the energy of the electron, and the curvature of the magnetic field lines. The relation between the particle energy spectral index ( $\alpha$ ) and the radiation spectral index ( $\beta$ ) is given as  $\alpha = 1 - 3\beta$  for the curvature radiation instead of  $\alpha = 1 - 2\beta$  for the synchrotron radiation. This type of production process for the VHE  $\gamma$  rays is expected to take place in pulsars and supernova remnants.

- **Bremsstrahlung**

Acceleration/de-acceleration of charged particles in electric fields is another production mechanism of  $\gamma$ -rays. When an electron passes by a positively charged nucleus, the trajectory of the electron is altered leading to emission of electromagnetic radiation. This process is known as bremsstrahlung. If the parent electrons have an energy  $N(E_e) \sim E^{-\alpha}$ , then the typical spectrum for bremsstrahlung is given as  $N(E_\gamma) \sim E_\gamma^{-\beta}$  where  $\alpha = \beta$ . The frequency range of this radiation depends on how much the electron trajectories are bent by the interaction with the positive ions or nucleus. This depends on the relative velocities of the two bodies, which in turn depends on the temperature of the gas. An example of high-energy thermal bremsstrahlung is the X-ray emission from giant elliptical galaxies and hot inter-cluster gas. The high-energy thermal bremsstrahlung does play a very important role in studies of diffuse Galactic emission for energies less than 200 GeV, but it is not a primary TeV  $\gamma$ -ray production mechanism in supernova remnants and pulsars.

- **Inverse Compton Scattering**

If photons of lower energy collide with energetic electrons, they gain energy in the collisions. This process is known as the inverse Compton (IC) process, illustrated in Fig A.1. The cross section of IC-scattering is approximately described by Thomson scattering, only when the photon's energy in the electron rest frame is less than the electron mass ( $E_\gamma \ll m_e c^2$ ). It is given as follows

$$\sigma_T = \frac{8\pi}{3} r_e^2 \quad (\text{A.4})$$

where  $r_e = 2.8 \times 10^{-13}$  cm is the classical electron radius. However, around MeV energies, where  $E_\gamma \geq m_e c^2$ , the cross section of the interaction is described by the Klein-Nishina formula, which is as follows:

$$\sigma_{KN} = r_e^2 \frac{\pi m_e c^2}{E_\gamma} \left[ \ln \left( \frac{2E_\gamma}{m_e c^2} + \frac{1}{2} \right) \right] \quad (\text{A.5})$$

Inverse Compton scattering is considered from relativistic electrons, which have a power-law distribution of the form  $N(E_e) \sim E^{-\alpha}$ , with a soft photon density of  $\rho_p h$ . The resulting  $\gamma$ -rays will have characteristic (Thompson) energies of  $\Gamma^2 h\nu$  or (Klein-Nishina) energies of  $\Gamma h\nu$ , where  $h\nu$  is the energy of the soft photon and  $\Gamma$  is the Lorentz factor. The spectrum of the resulting  $\gamma$ -rays is given as  $N_\gamma(E) \sim E_\gamma^{-(\alpha+1)/2}$ . This process is important in regions with high photon densities. For example, in compact objects like neutron stars, which generate beams of charged particles in their vicinity.

- **Annihilation in flight**

The cross section for  $e^+e^-$  annihilation at extreme relativistic energies is given by

$$\sigma_A = \frac{\pi r_e^2}{\gamma} [\ln(2\gamma) - 1] \quad (\text{A.6})$$

where  $\gamma(= E_e/m_e c^2)$  is the Lorentz factor of the positrons and  $r_e$  is the classical radius of the electron. Two photon production is the dominant result of annihilation leading to a continuum of photon energies. At high positron energies, the photon emitted in the extreme forward direction has almost all the energy of the positron while the other photon has an energy of  $m_e c^2/2 = 0.256 \text{ MeV}$ . The cross section given by the above equation is quite low ( $\sigma_A = 6 \times 10^{-27} \text{ cm}^2$  for a 100 MeV positron).

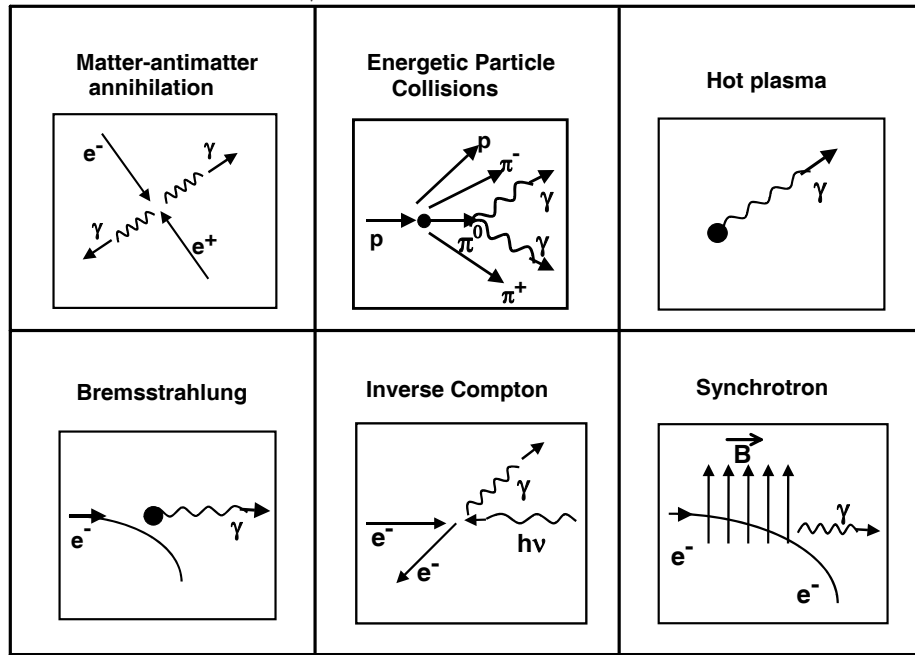


Figure A.1 : Pictorial representation of the various  $\gamma$ -ray production mechanisms.

- **Pair annihilation**

Annihilation between particles and antiparticles may also produce  $\gamma$ - rays. The annihilation process of an electron and positron is also shown in Fig A.1. The rays from pair annihilation are produced in the vicinity of radioactive decay regions, or energetic environments capable of positron production by other processes. Examples of the pair annihilation can be seen in neutron stars

or black holes, which have high energy densities and strong gravitational and magnetic fields. In a similar manner, hadronic anti-particles also annihilate with their corresponding particles. This may cause the spectral features at correspondingly higher energies in the  $\gamma$ -ray spectrum.

- **Pion Production and Decay**

Pion is an elementary particle, which is created in strong interaction processes, such as a collision of an accelerated cosmic-ray proton with a nucleus in the ambient gas or with another proton or high-energy photon. In such an interaction charged ( $\pi^\pm$ ) or neutral ( $\pi^0$ ) secondary pions are created. The neutral pions then rapidly decay into two rays, each of which has an energy of  $\sim 70$  MeV in the rest frame of  $\pi^0$ . The charged pions decay into muons and neutrinos. If the cosmic-ray has a power-law spectrum with a spectral index of  $\alpha$ , then for higher energies the  $\gamma$ -ray spectral distribution is a power-law with spectral index  $\beta = (4/3)(\alpha - 1/2)$ . When the energy decreases, the spectrum turns over with a peak at 70 MeV. This peak is the characteristic feature of the p-p interaction and a signature of hadrons as primary cosmic particles.

# Appendix B

## Interaction of $\gamma$ -rays

Though a large number of interaction mechanisms are known for  $\gamma$ -rays, only three major types play an important role in radiation measurements namely the photoelectric absorption, Compton scattering and pair production.

- **Photoelectric absorption**

In this process, a photon undergoes interaction with an absorber atom in which the photon completely disappears. In its place an energetic photoelectron is ejected by the atom from one of its bound shells. The interaction is with the atom as a whole and cannot take place with free electrons. For  $\gamma$ -rays of sufficient energy, the most probable origin of photoelectron is the most tightly bound or  $K$  shell of the atom. The photoelectron appears with an energy given by

$$E_e = h\nu - E_b \quad (\text{A.7})$$

where  $E_b$  represents the binding energy of the photoelectron in its original shell. For  $\gamma$ -ray energies of more than a hundred keV, the photoelectron carries off the majority of the original photon energy. In addition to the photoelectron, the interaction also creates an ionized absorber atom with a vacancy in one of its bound shells. This vacancy is quickly filled through capture of a free electron from the medium and /or rearrangement of electrons from other shells of the atom. Therefore, one or more characteristic X-ray photons may also be generated. Although in most cases these X-rays are reabsorbed close to the original site through photoelectric absorption involving less tightly bound shells, their migration and possible escape from detectors can influence their response.

- **Compton scattering**

The interaction process of compton scattering takes place between the incident  $\gamma$ -ray photon and an electron in the absorbing material. It is most often the predominant interaction mechanism for  $\gamma$ -ray energies typical of radioisotope sources.

In this type of scattering, the incoming  $\gamma$ -ray photon is deflected through an angle  $\theta$  with respect to the original direction. The photon transfers a portion of its energy to the electron (assumed to be initially at rest), which is then known as recoil electron. Because all angles of scattering are possible, the energy transferred to the electron can vary from zero to a large fraction of the  $\gamma$ -ray energy. The photon gets scattered with the energy given as

$$h\nu' = \frac{h\nu}{1 + \frac{h\nu}{m_0c^2}(1 - \cos\theta)} \quad (\text{A.8})$$

where  $m_0c^2$  is the rest mass energy of the electron. For small scattering angle  $\theta$ , very little energy is transferred. Some of the energy is always retained by incident photon, even in the extreme case of  $\theta = \pi$ . The probability of the Compton scattering per atom of absorber depends on the number of electrons available as scattering targets and thus increases linearly with  $Z$

- **Pair production**

If the  $\gamma$ -ray energy exceeds the rest mass energy of the electron, the process of pair production is energetically possible. The probability of this interaction remains very low until the  $\gamma$ -ray energy approaches several MeV and therefore pair production is predominantly confined to high energy  $\gamma$ -rays. In such an interaction, which takes place in the coulomb field of a nucleus, the  $\gamma$ -ray photon disappears and is replaced by an electron-positron pair. All the excess energy carried by the photon above 1.02 MeV required to create the pair goes into kinetic energy shared by the positron and the electron. Because the positron will subsequently annihilate after slowing down in the absorbing medium, two annihilation photons are normally produced as secondary products of the interaction. Though no simple rule exists for the probability of pair production per nucleus, its magnitude however varies approximately as the square of the absorber atomic number.

The relative importance of the three processes for different absorber materials and  $\gamma$ -ray energies is illustrated in the Fig B.1. The curve on the left represents the energy, at which photoelectric absorption and compton scattering are equally probable as a function of the absorber atomic number. The curve on the right represents the energy at which compton scattering and pair production are equally probable. Three areas are thus defined on the plot within which the photoelectric absorption, Compton scattering and the pair production mechanisms dominate.

- **$\gamma - \gamma$  interaction**

This process is of extreme importance and requires very unusual combination of high energy photons and a high density of low energy photons.  $\gamma$ -rays are

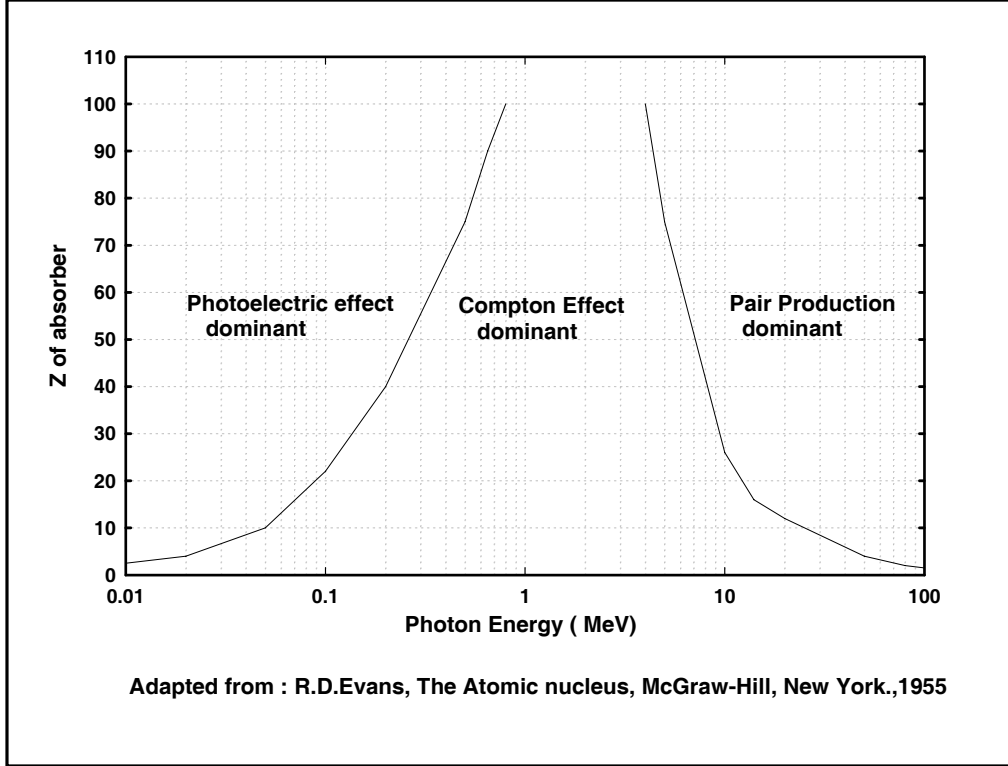


Figure B.1: Gamma-ray interaction as a function of  $\gamma$ -ray production mechanisms.

absorbed by photon-photon pair production ( $\gamma + \gamma \rightarrow e^+ + e^-$ ) on background photon fields provided the center of mass energy of the two photon system is more than the square of rest mass energy of an electron. A  $\gamma$ -ray photon of energy  $\epsilon_1$  collides with another photon of lower energy  $\epsilon_2$  and gives rise to a pair of particles, each of mass  $m$  if  $\epsilon_1$  is greater than a threshold value  $\epsilon_t$ , given by :

$$\epsilon_t = \frac{2m^2 c^4}{\epsilon_2(1 - \cos\theta)} \quad (\text{A.9})$$

where  $\theta$  is the angle between the photon trajectories. Taking a case of head on collisions,  $\epsilon_2 \epsilon_t = 0.26 \times 10^{12}$  (all energies in eV). Following are the cases with astronomically important radiation fields with their mean energy ( $\epsilon_m$ ) and required threshold energy of incident photon: (i)  $2.7K$  CMBR ( $\epsilon_m \sim 6 \times 10^{-4} \text{eV}$ ),  $\epsilon_t \sim 4 \times 10^{14} \text{eV}$ , (ii) Starlight ( $\epsilon_m \sim 2 \text{eV}$ ),  $\epsilon_t \sim 10^{11} \text{eV}$  (iii) X-ray ( $\epsilon_m \sim 1 \text{KeV}$ ),  $\epsilon_t \sim 3 \times 10^8 \text{eV}$ .

The cross section for this process is not very small but such collisions are not very frequent because of the low densities of target photons. Nevertheless, even with low photon densities these collisions lead to significant attenuation of photons travelling over typical extragalactic distances. The effect is par-



ticularly important where the presence of EBL, limits the distance to which VHE  $\gamma$ -ray telescopes can detect sources.