

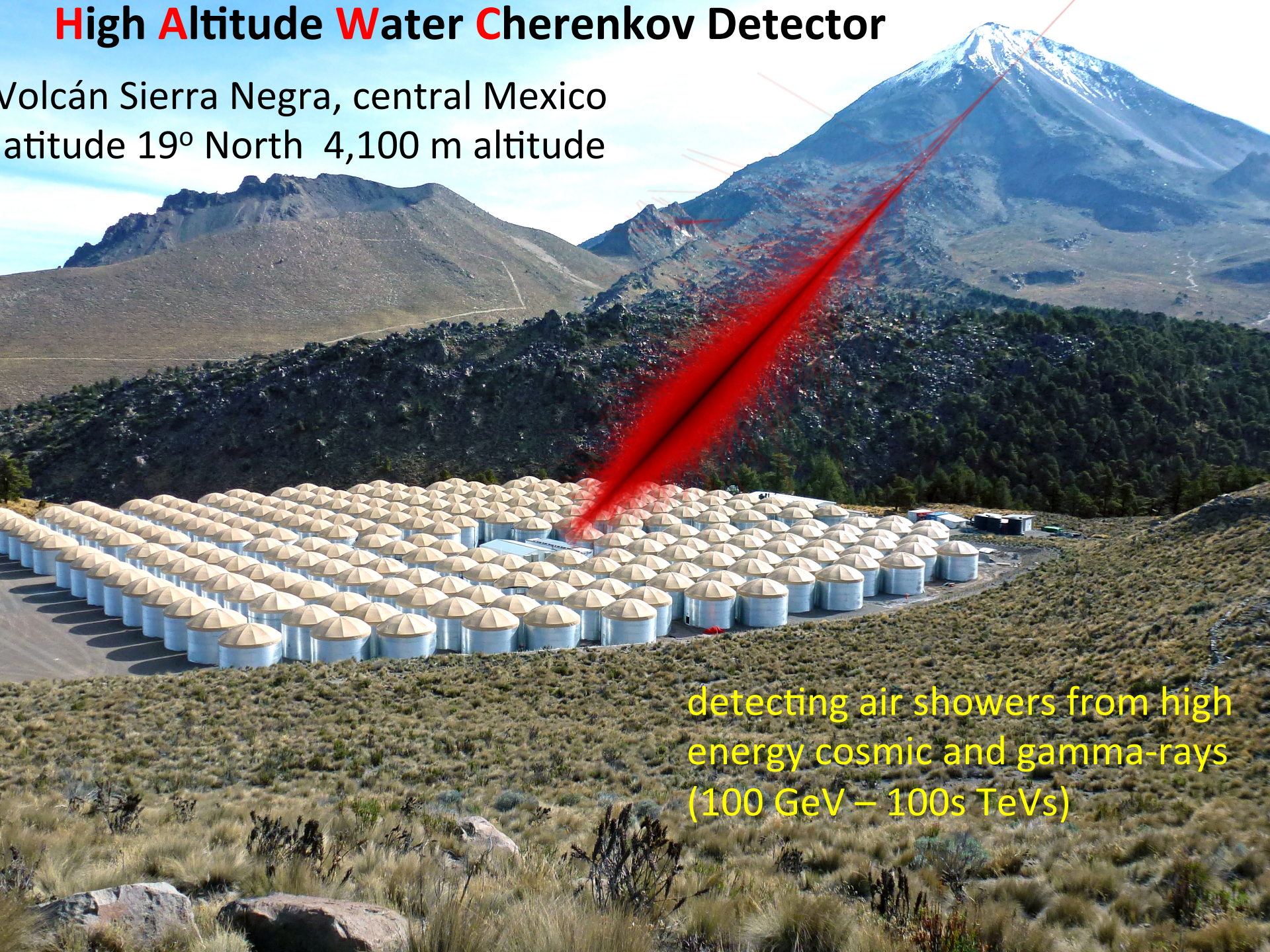
# Data storage and analysis of the HAWC gamma ray observatory

Andres Sandoval  
asandoval@fisica.unam.mx  
Instituto de Fisica, UNAM  
23 de abril 2015



# High Altitude Water Cherenkov Detector

Volcán Sierra Negra, central Mexico  
latitude 19° North 4,100 m altitude



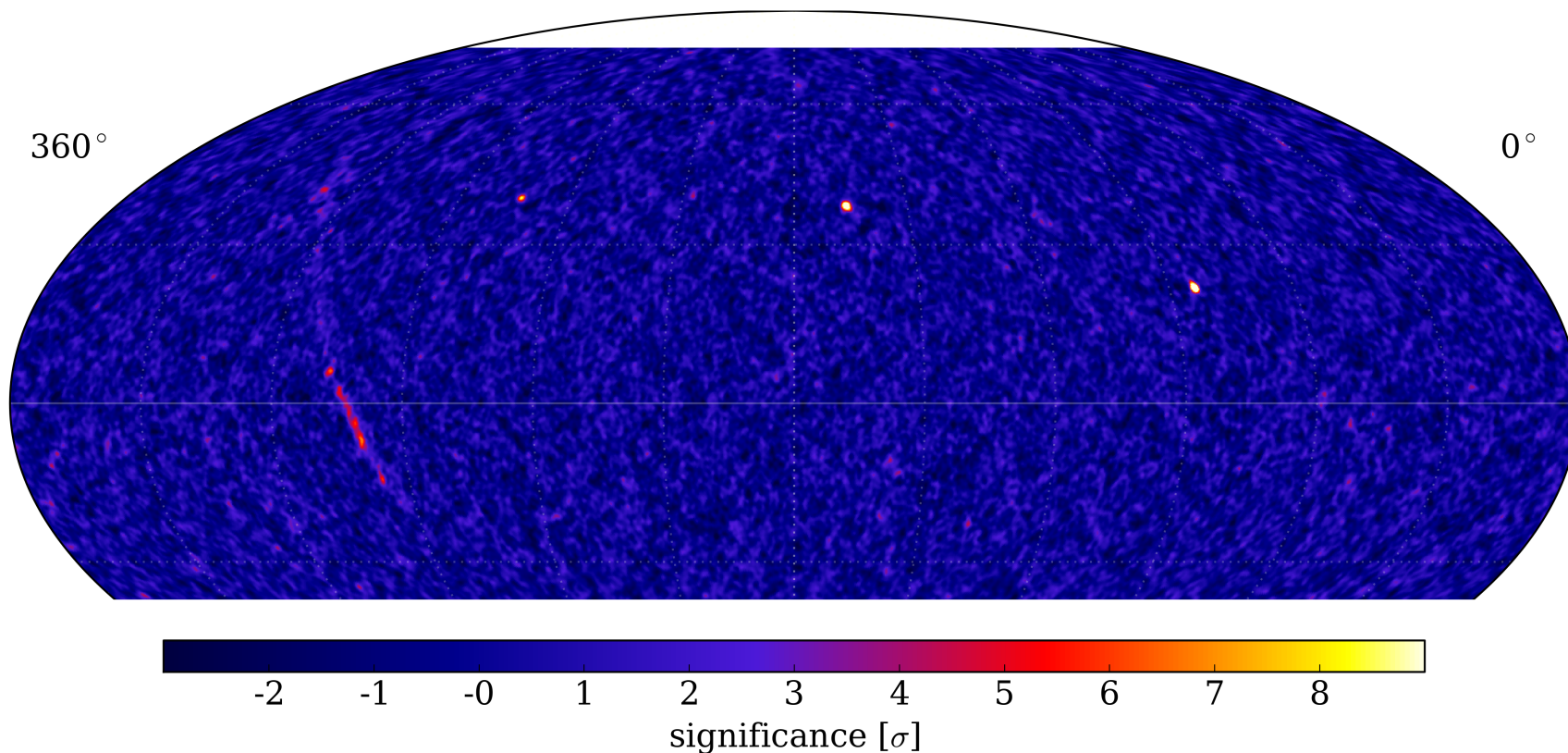
detecting air showers from high  
energy cosmic and gamma-rays  
(100 GeV – 100s TeVs)

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.

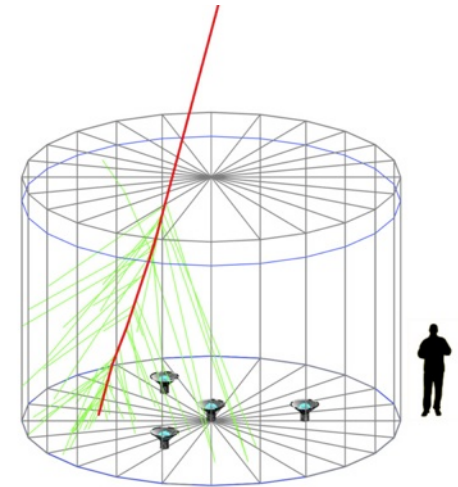
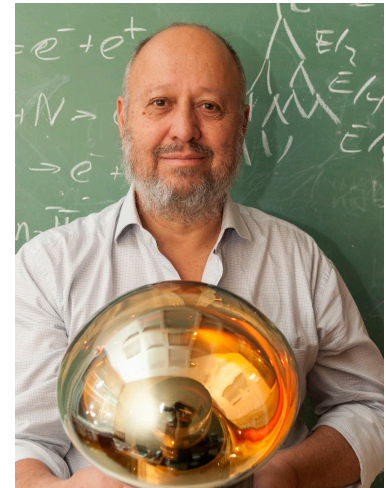
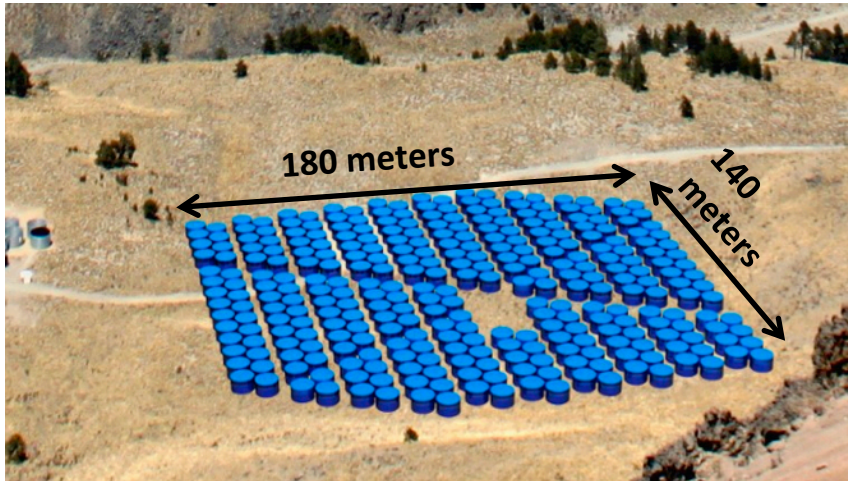


Make sky maps in  
 $\gamma$  rays day by day for 10 years

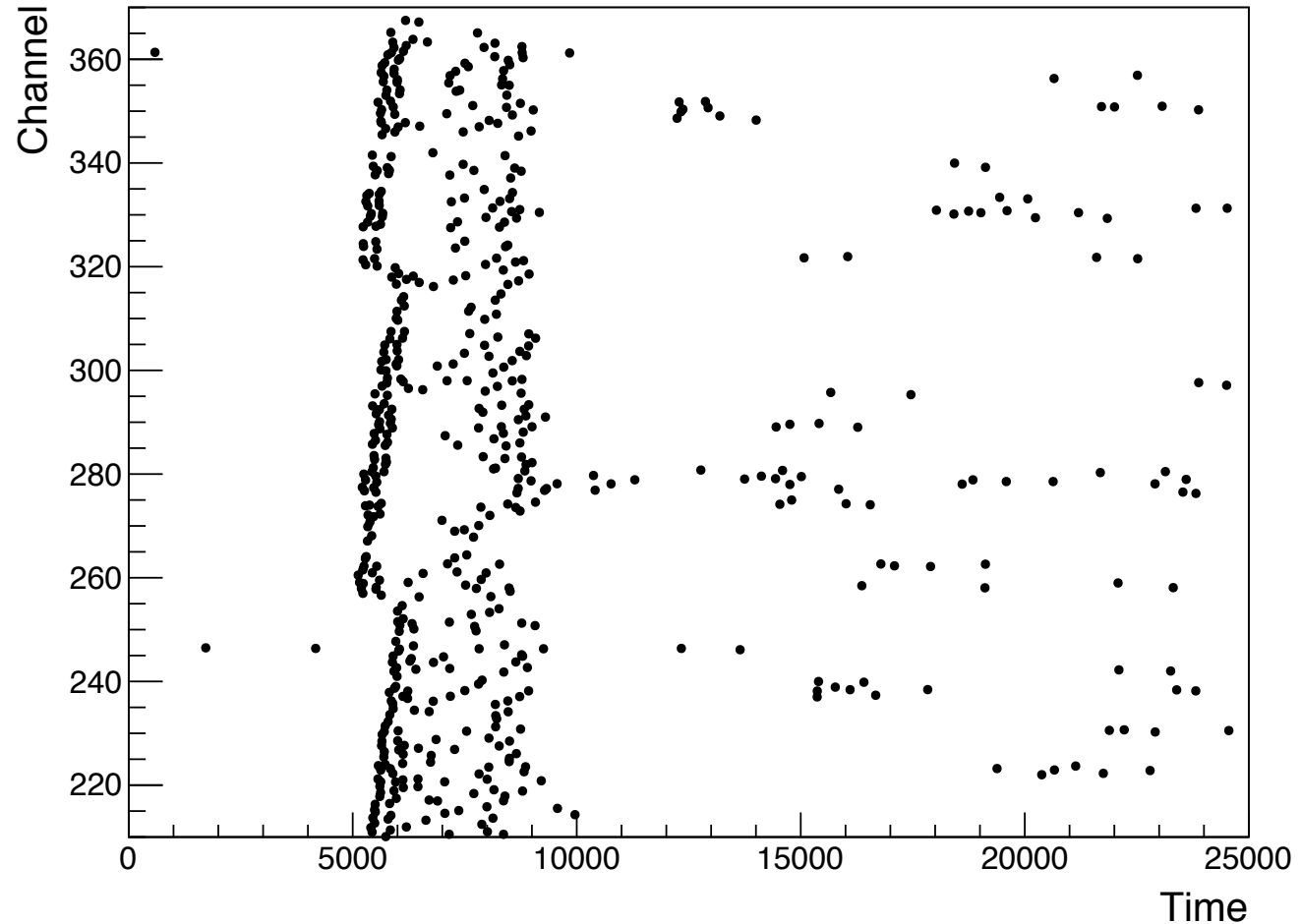


# HAWC Design

300 close packed water tanks (7.3m dia x 4.5 m deep of 200,000 liters) each with 4 upward facing photomultiplier tubes at the bottom

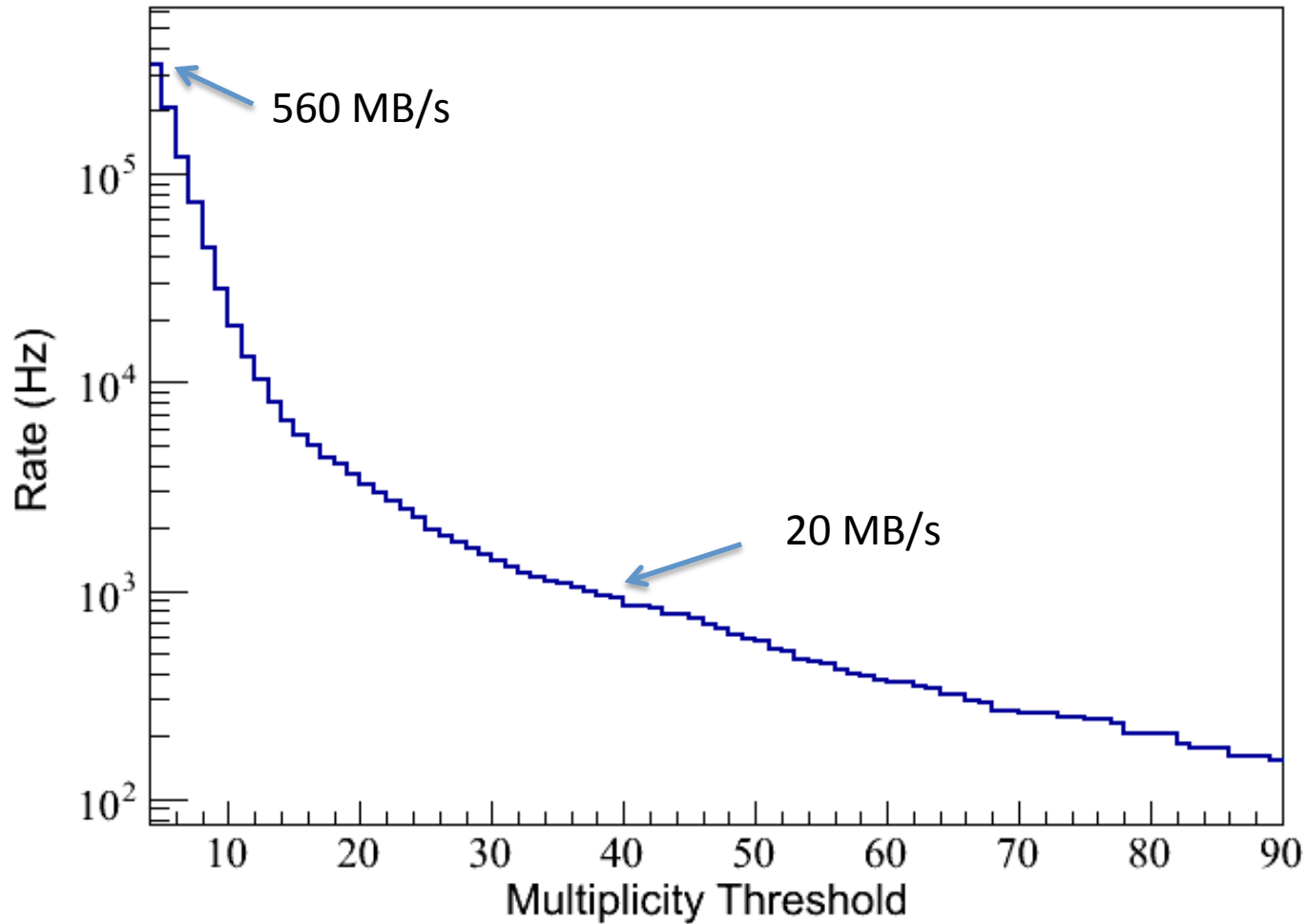


# Signals of the 1,200 detectors are digitized in 100 picosecond time bins





# Event rate vrs multiplicity





# Computing Strategy

- Two analysis and data centers
  - Instituto de Ciencias Nucleares, UNAM
  - University of Maryland
- Full replication of raw- simulated and analyzed data
  - 1 PB per year at each site
- Data storage on disks
- Computing farms on commodity processors
- Able to process 1 year of data in 1 month



# Computing strategy

- Operating system Linux Centos
- languages C++, Python
- databases mysql
- packages: CORSICA, ROOT, GEANT, HEALpix
- wiki / SVN / TracBrowser / Doxygen/
- SeeVough for conferencing





# UMD Jan 2015

## Disks: (Gluster)

- /data/archive - 625TB (597TB used)
  - Contains all data up to 6/2014 + Milagro
- /data/archive2 - 397TB (182 TB used)
  - Contains data from 7/2014...
- /data/scratch - 49TB
- Computers: 13x48 core + 2x32 core = 688 cores



# UMD April 2015

Disks: (Lustre) 1.76 PB -> 2 PB this summer

- /hawc01: 675TB -> 1 PB this summer
- /hawc0: 640TB
- /archive2: 397 TB
- /data/scratch - 49TB
- Computers: 19x48 core + 2x32 core = 976 cores



# ICN

Lustre file system

January 2015

- 495 TB disk
- 4 x 48 cores

May 2015

- 1.36 TB
- 12 x 48 cores = 576 cores

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.



## Some hardware at UMD and ICN





Purchased November 2014, to  
be delivered May 2015

- 2x MDS (1x E5-2680v3, 64GB Memory, 1x 120GB SSD) mirrored storage servers
- 8x Compute Nodes (2x E5-2690v3, 128GB Memory, 2x 1TB SATA HDD) having 384 cores
- 6x JBOD (24x 6TB SAS HDD) 864 TB disk storage
- 108 Port Mellanox FDR Switch
- 6x OSS (1x E5-2680v3, 128GB Memory, 1x 120GB SSD)
- Parts (Switch, Cables, Drives)
- 1x JBOD (JBOD Kit and 12x 600GB SAS HDD)



## HAWC Site

Data acquisition ~20MB/s  
Copy to portable disk arrays  
(80 MB/s)  
Erase copied and verified  
data

Hand carried to  
UNAM

sent as packages:  
Estrella Roja

Check  
sums

## ICN

copied to permanent  
storage  
transfer to UMD  
(1 - 2 MB/s)

Data transfer  
1 GB/s network

## UMD

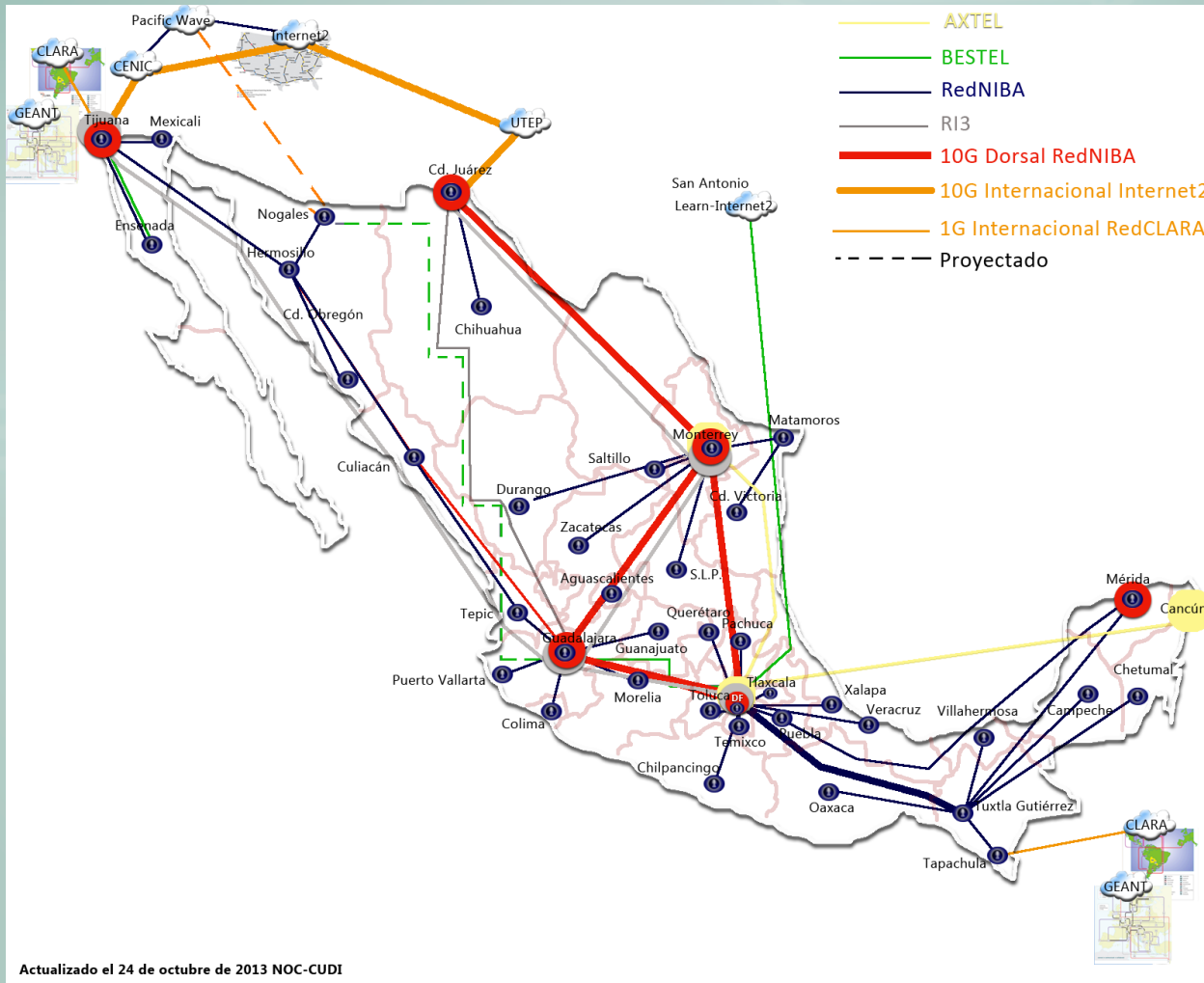
copied to  
permanent storage  
verify copy

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.



# UNAM – San Antonio



21 AL 24 DE ABRIL

Puerto Vallarta, Jal.



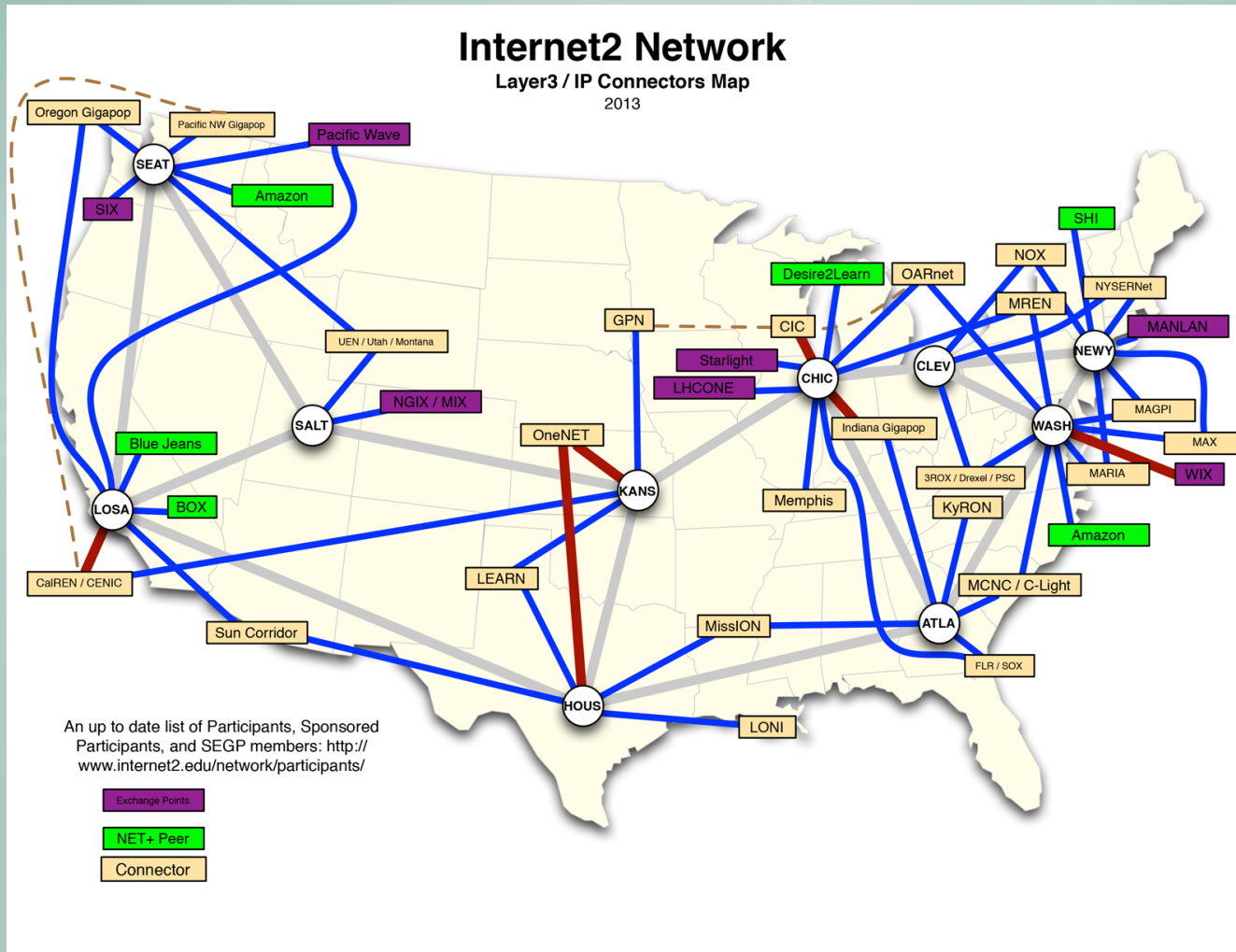
# San Antonio - Houston





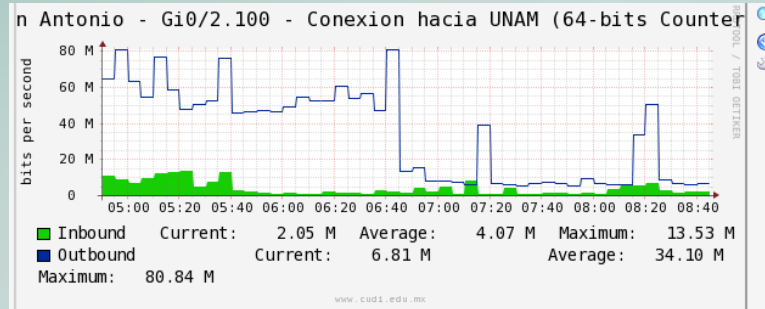


# Houston - Maryland

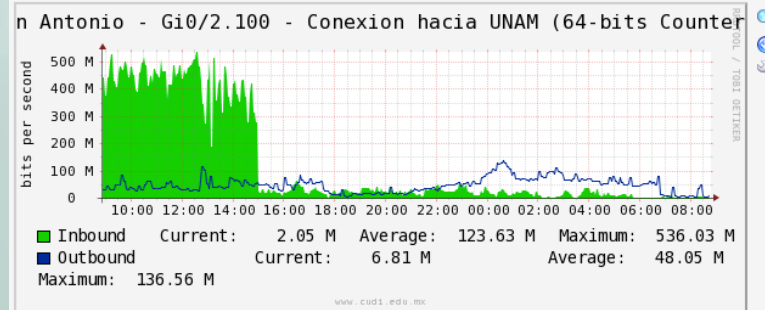




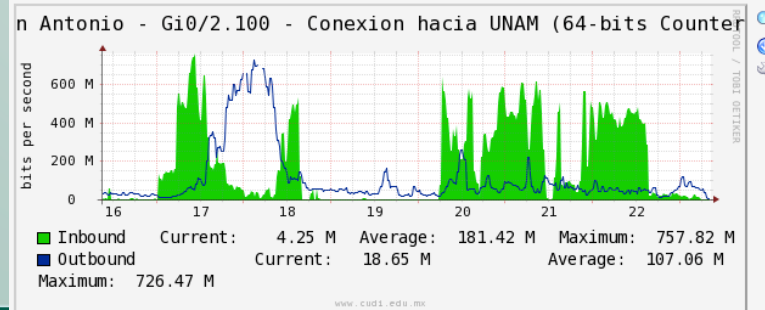
# Data transfer UNAM-San Antonio



Hourly (1 Minute Average)



Daily (5 Minute Average)

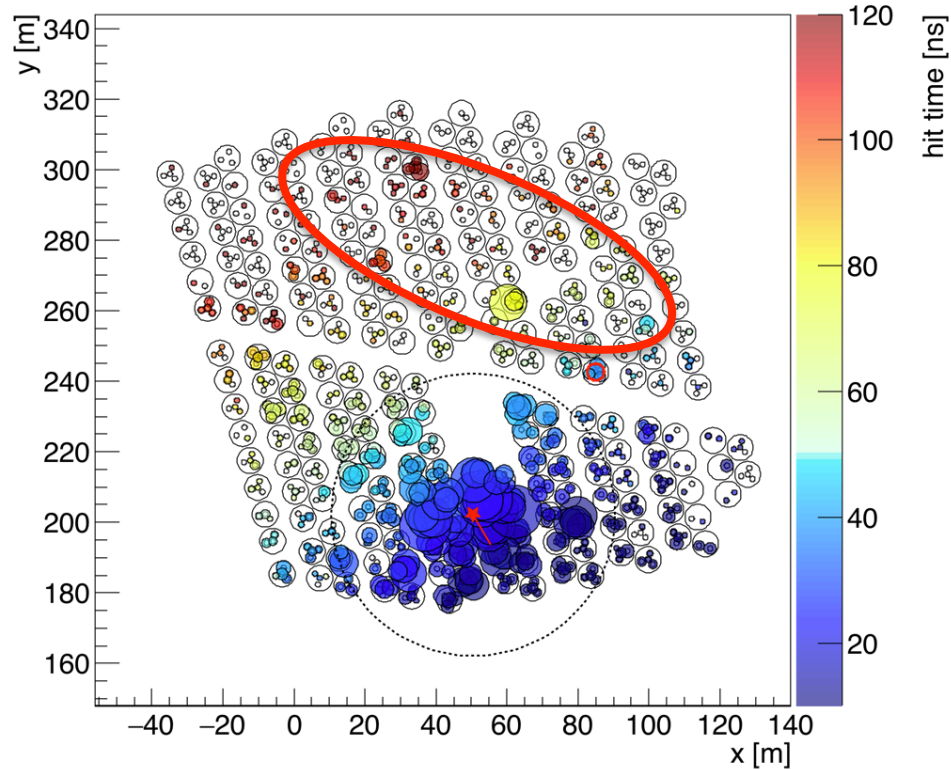


Weekly (30 Minute Average)

# HAWC-250 Data

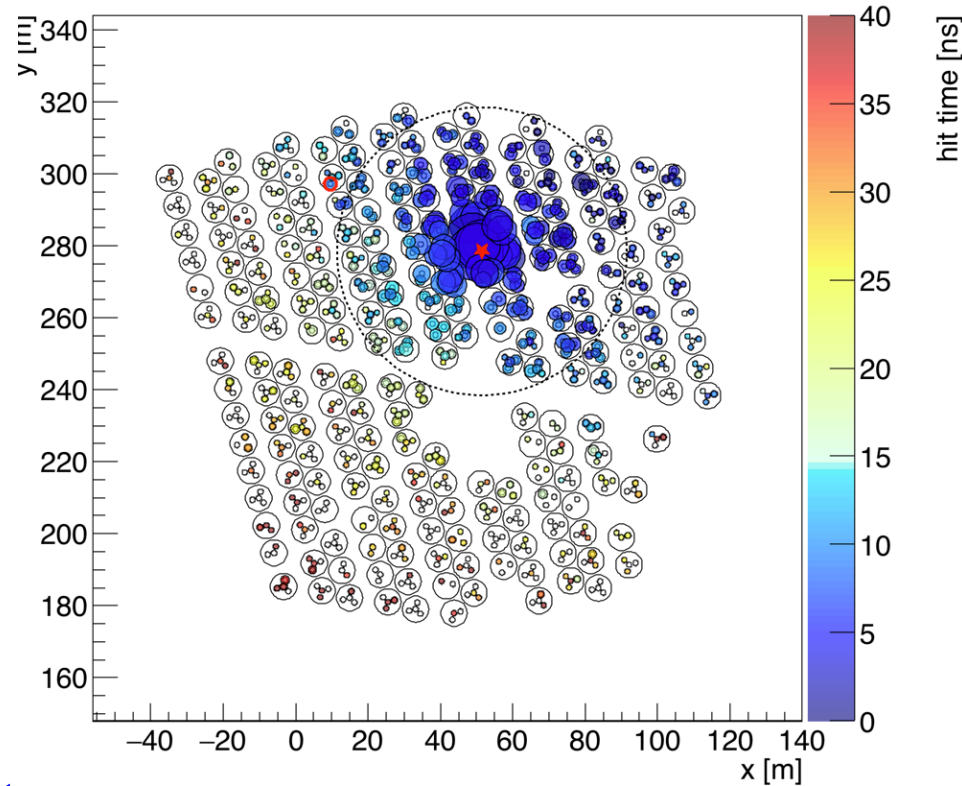
Run 2118, TS 45004, Ev# 41, CXPE40= 55.7, Cmptness= 10.7

Run 2054, TS 584212, Ev# 226, CXPE40= 21.2, Cmptness= 28.3



Hadron Shower  
(off source)

## Data

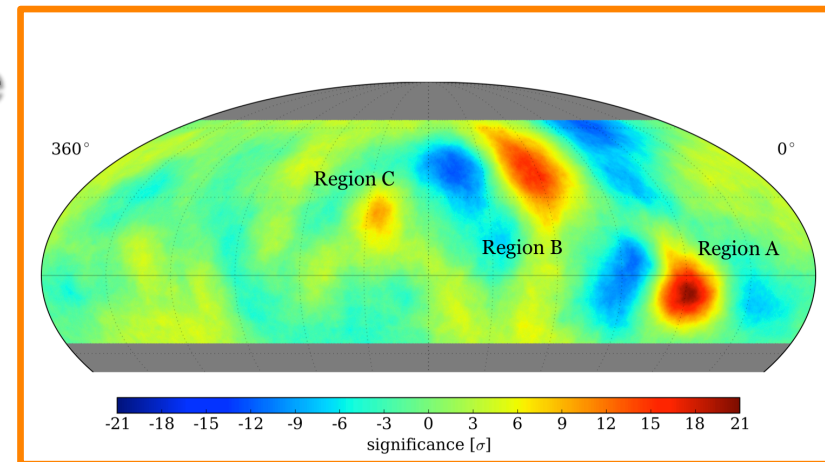
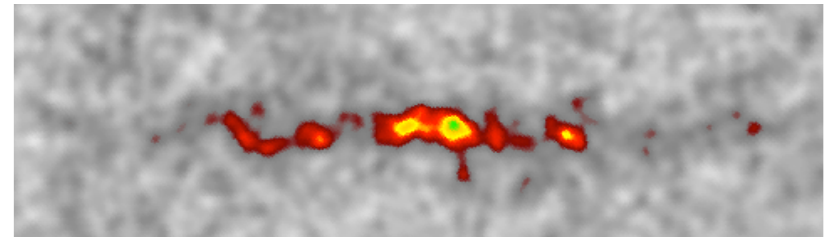
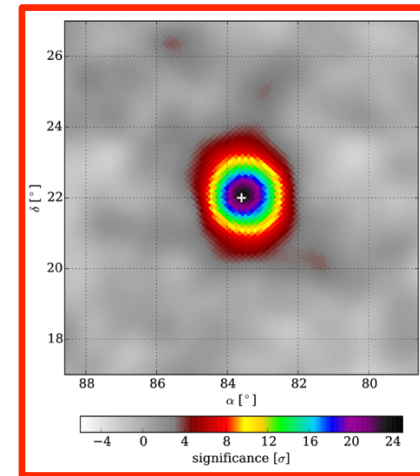


Likely Gamma Shower  
(Crab event)



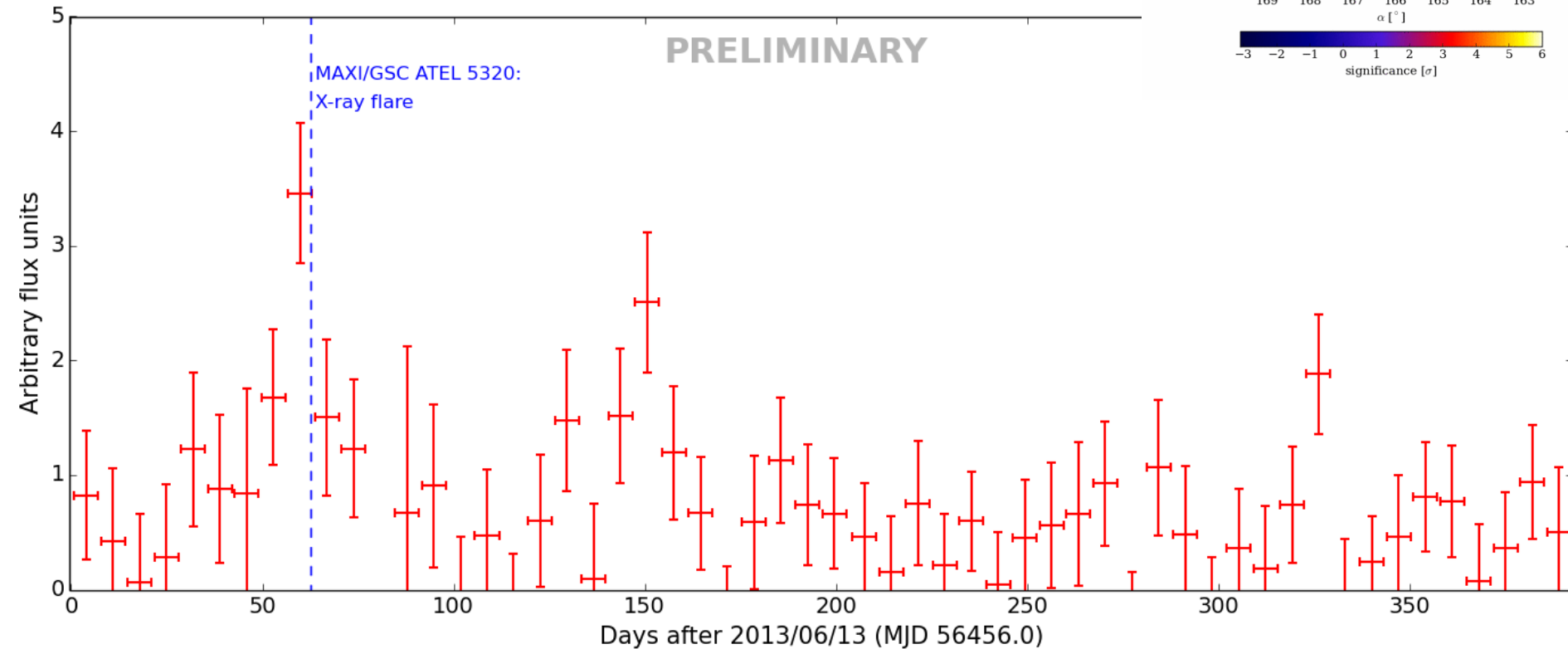
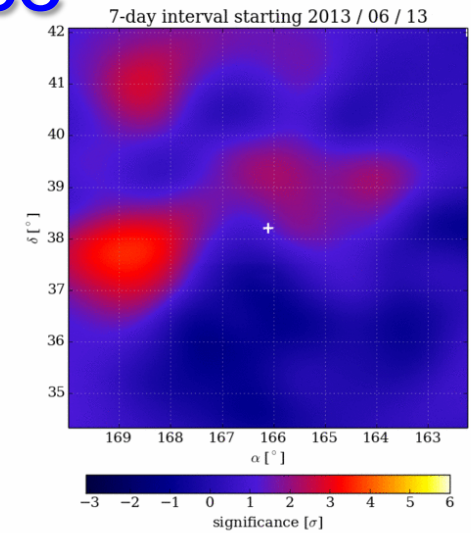
# HAWC Science

- **Gamma Ray Astrophysics**
  - Galactic Gamma-Ray Sources
    - Pulsars, PWNs, Binaries
    - Galactic Diffuse and Fermi Bubbles
  - Extragalactic Gamma-Ray Sources
    - Gamma-ray bursts
    - Active Galactic Nuclei
- **Particle Physics**
  - Dark Matter
  - Primordial Black Holes
  - Violations of Lorentz Invariance
  - Q-balls
- **Cosmic Ray Anisotropy**
- **Solar Physics**
  - CMEs



# Mrk 421 Time Dependence

## Mrk 421 HAWC-111





# Event reconstruction is done centrally

## Reconstruction v2.00.03

Task currently handling 28768 files totaling 40892.36 GB:

WAITING: 0

PROCESSING: 0

RECONSTRUCTED: 28768

Processing time invested in reconstructed files only: 126236.96 core-hours

Here are the 90 fully reconstructed runs for this task:

2250 2251 2252 2256 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268  
2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282  
2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296  
2297 2298 2299 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311  
2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325  
2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2339 2340  
2341 2343 2344 2345 2346 2347



# Standard reconstruction

Hawcprod task ID	Run range	Start date	End date	Data size	Core-hours	Rate
66	2179-2197	22 Dec.	3 Jan.	12.5 TB	31854	392 MB/hr
63	2166-2178	16 Dec.	22 Dec.	6.6 TB	15345	430 MB/hr
60	2148-2165	10 Dec.	16 Dec.	6.5 TB	15177	430 MB/hr
56	2107-2147	29 Nov.	10 Dec.	13.5 TB	35989	375 MB/hr

Month of data amounts to ~39 TB of data (1 week on 600 cores)

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.



Event simulation is done centrally  
in Deep Thought 2 at UMD



8,000 cores  
1 PB storage  
10 GB/s link





## Huge effort to simulate and reconstruct shower events

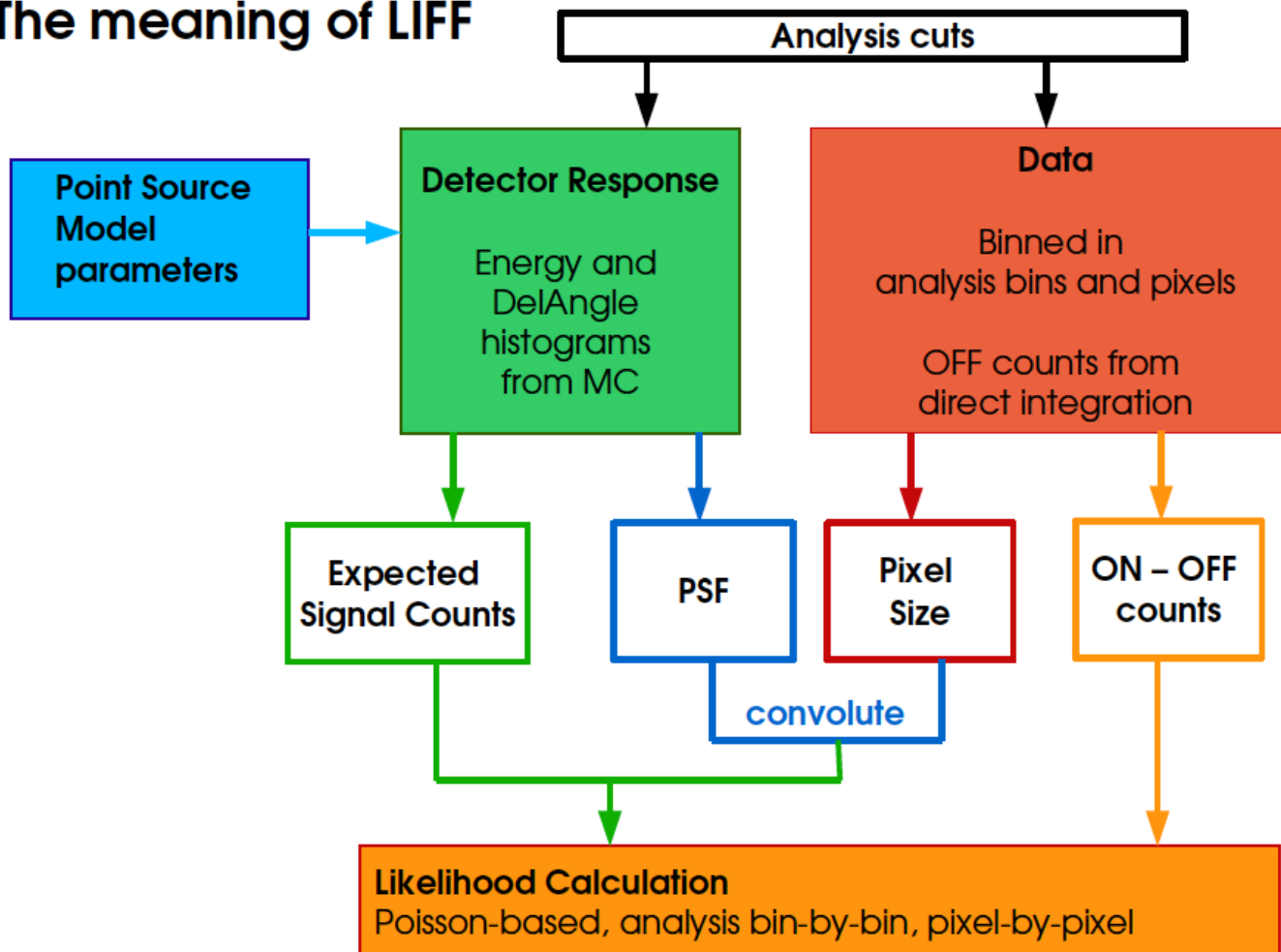
### New Production

PRIMARY	RUNS	EVENTS/ RUN	TOTAL EVENTS
Gamma	300000	100000	3E+10
Proton	300000	100000	3E+10
Helium	130000	100000	1.3E+10
Carbon	20000	50000	1E+09
Oxygen	20000	50000	1E+09
Neon	20000	50000	1E+09
Magnesium	20000	50000	1E+09
Silicon	20000	50000	1E+09
Iron	20000	50000	1E+09



# Event analysis is done individually

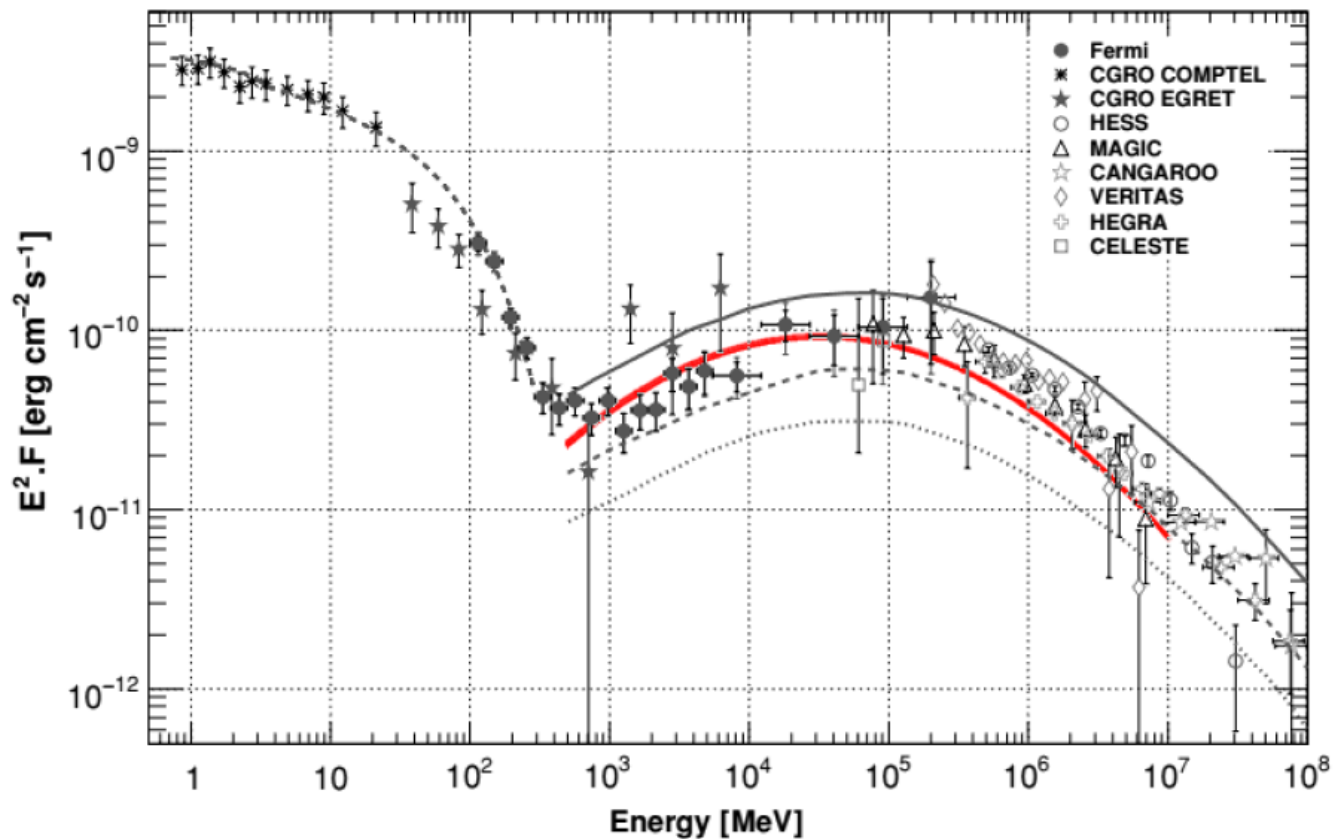
## The meaning of LIFF





Over 100 users will be doing  
analysis

## Joint Fermi-HAWC 3ML fit of the Crab spectrum



3ML fit of 29 days of Fermi-LAT and 30 days of HAWC data

# Cosmic Ray anisotropy

## Data Set

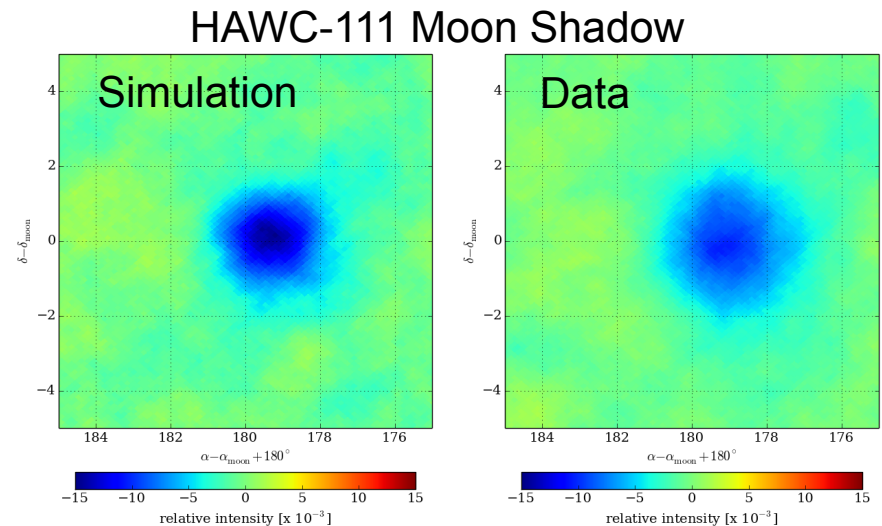
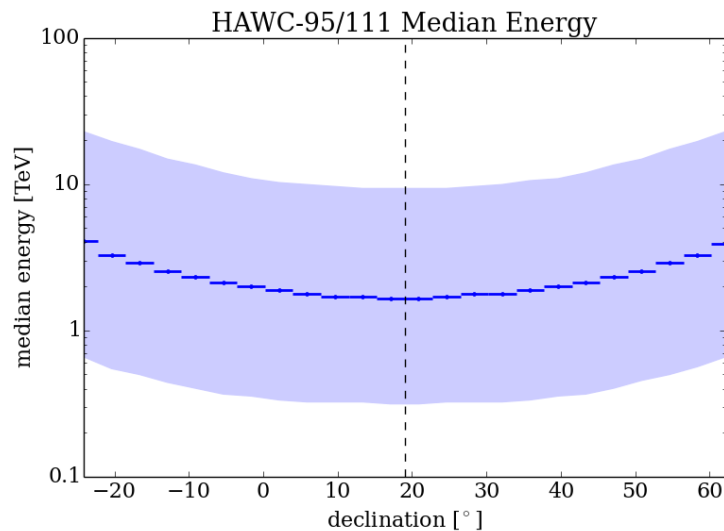


Using HAWC-95 and HAWC-111

June 2013 – February 2014

114 full sidereal days

50 billion events,  
1.2° median ang. res.,  
1.8 TeV median energy

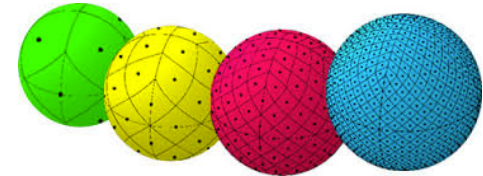


Observation of Small-scale Anisotropy in the Arrival Direction Distribution of TeV Cosmic Rays with HAWC [arXiv:1408.4805 \[astro-ph.HE\]](https://arxiv.org/abs/1408.4805)

# Analysis Technique

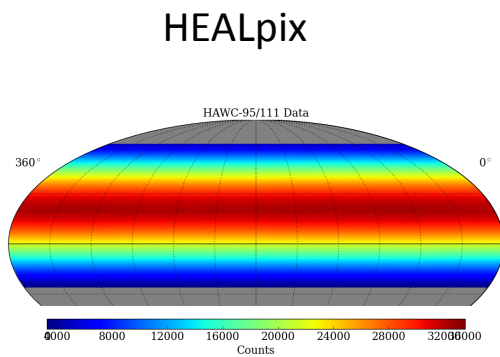


**HEALpix** (K.M. Gorski et al., *Astrophys. J.*, 2005, 622, 759)  
Equal-area binning of the sphere

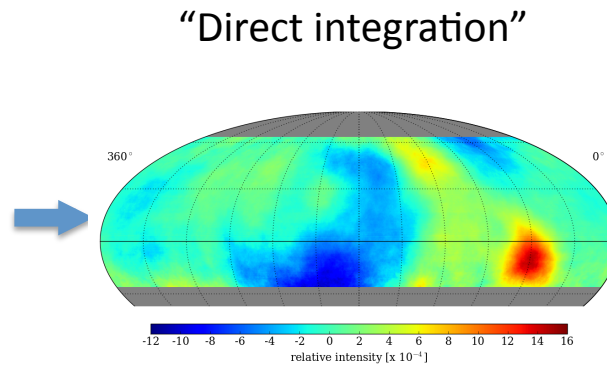


**“Direct Integration”** (R. Atkins et al., *Astrophys. J.*, 2003, 595, 803.)  
Method to estimate background using the data themselves

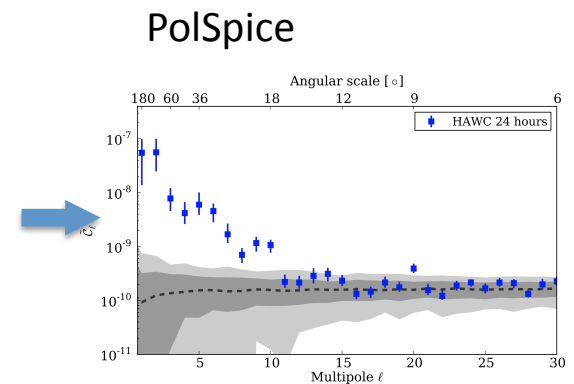
**PolSpice** (I. Szapudi et al. 2001, *Astrophys. J.*, 548, L115)  
Software to compute power spectrum with partial sky coverage



Binned data



Data & reference map  
→ relative intensity



Power Spectrum

21 AL 24 DE ABRIL

Puerto Vallarta, Jal.



HAWC will be in the Big Data  
Big Network game for the  
foreseeable future



21 AL 24 DE ABRIL

Puerto Vallarta, Jal.



# HAWC funding

