Computational issues for space experiments: AMS-02

D. D'Urso(INFN & ASI-ASDC)

Outline

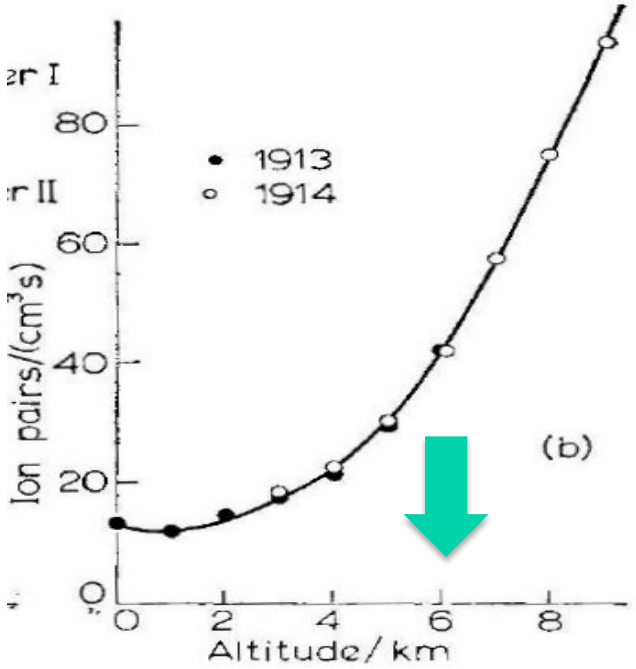
1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

Outline

1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

A long history of studies

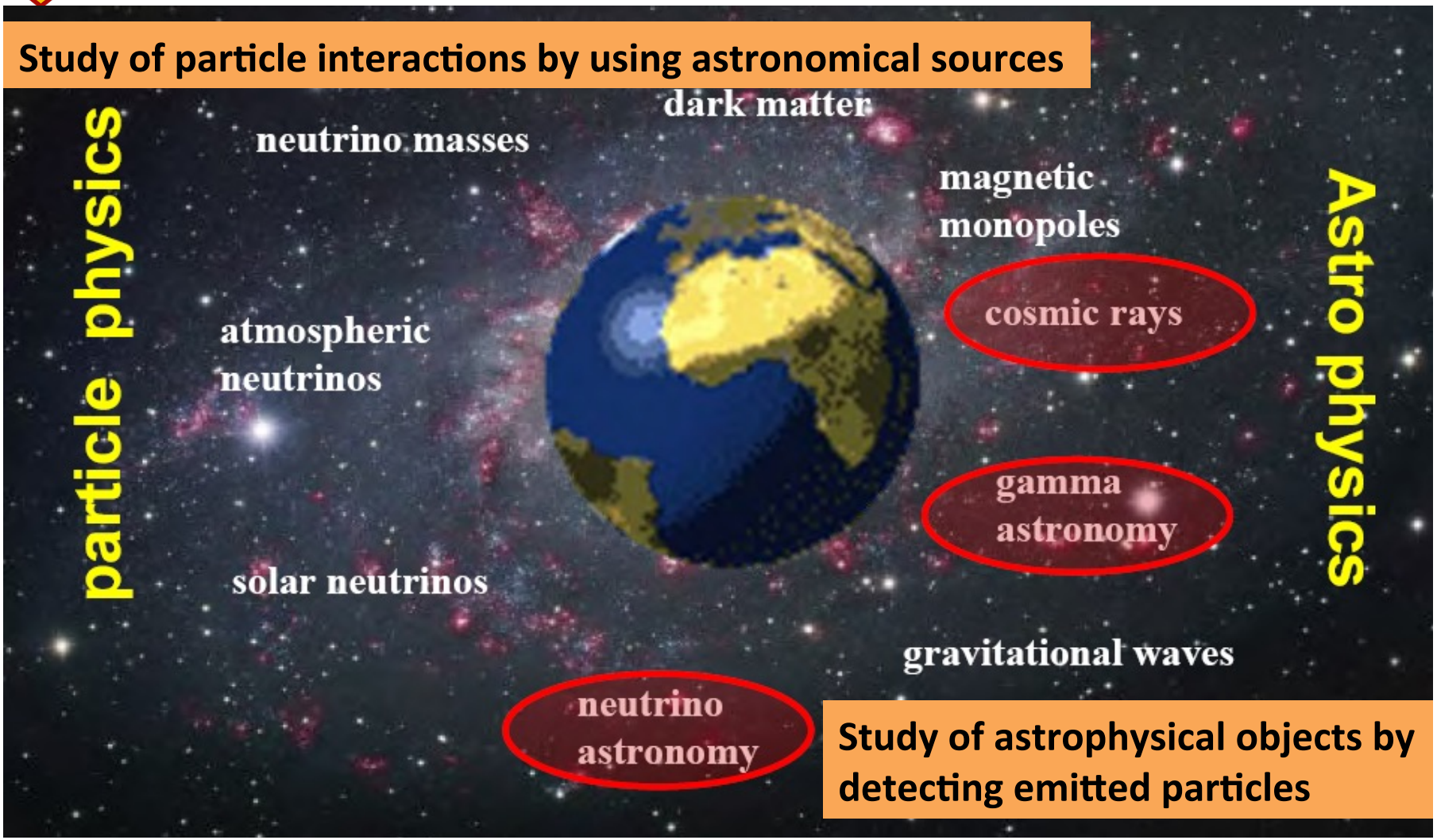
Viktor Hess (1910-1914)
several balloon flights to measure radiation as a function of height



Source outside the Earth

Astroparticle Physics

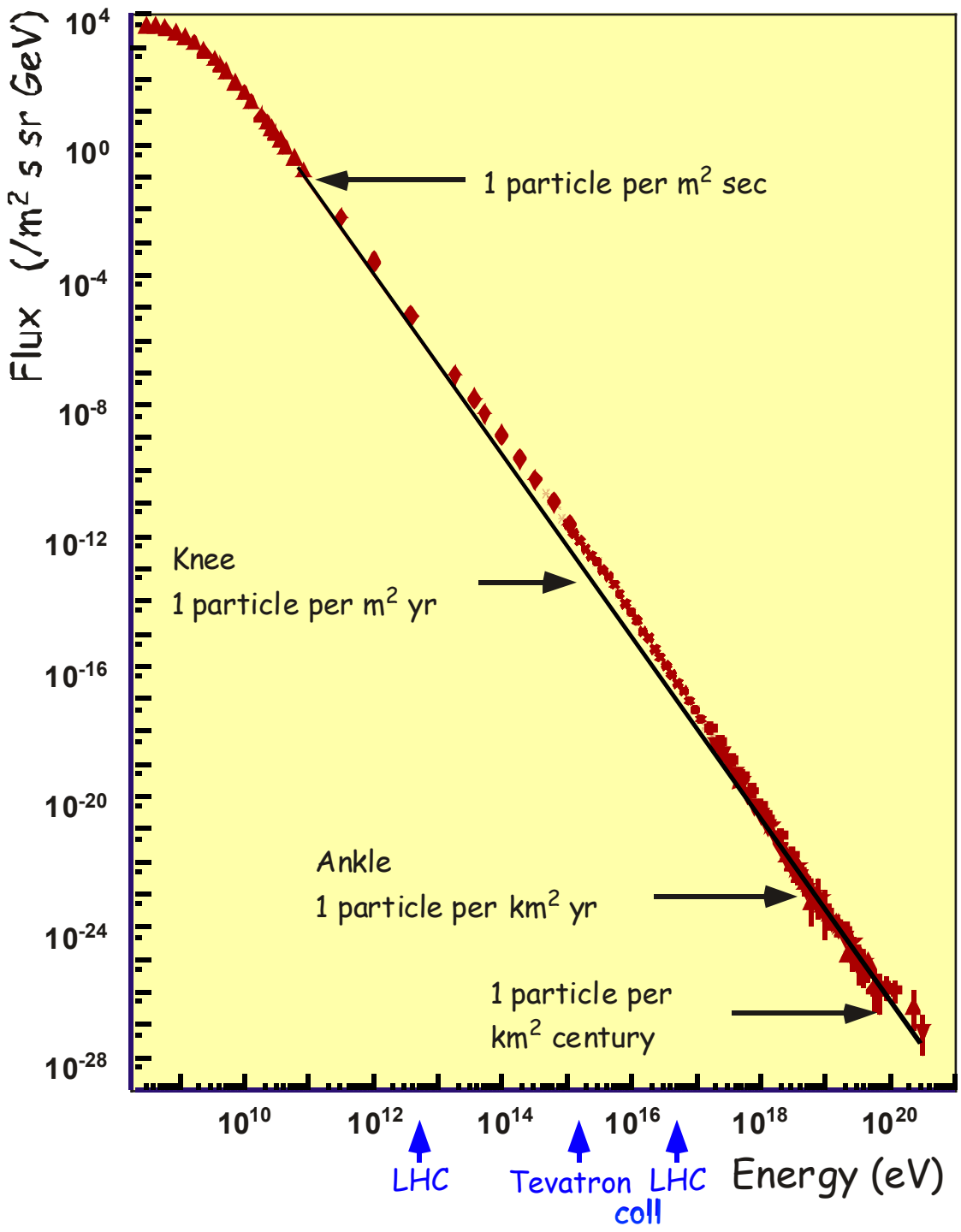
Astroparticle Physics



Why to study CRs

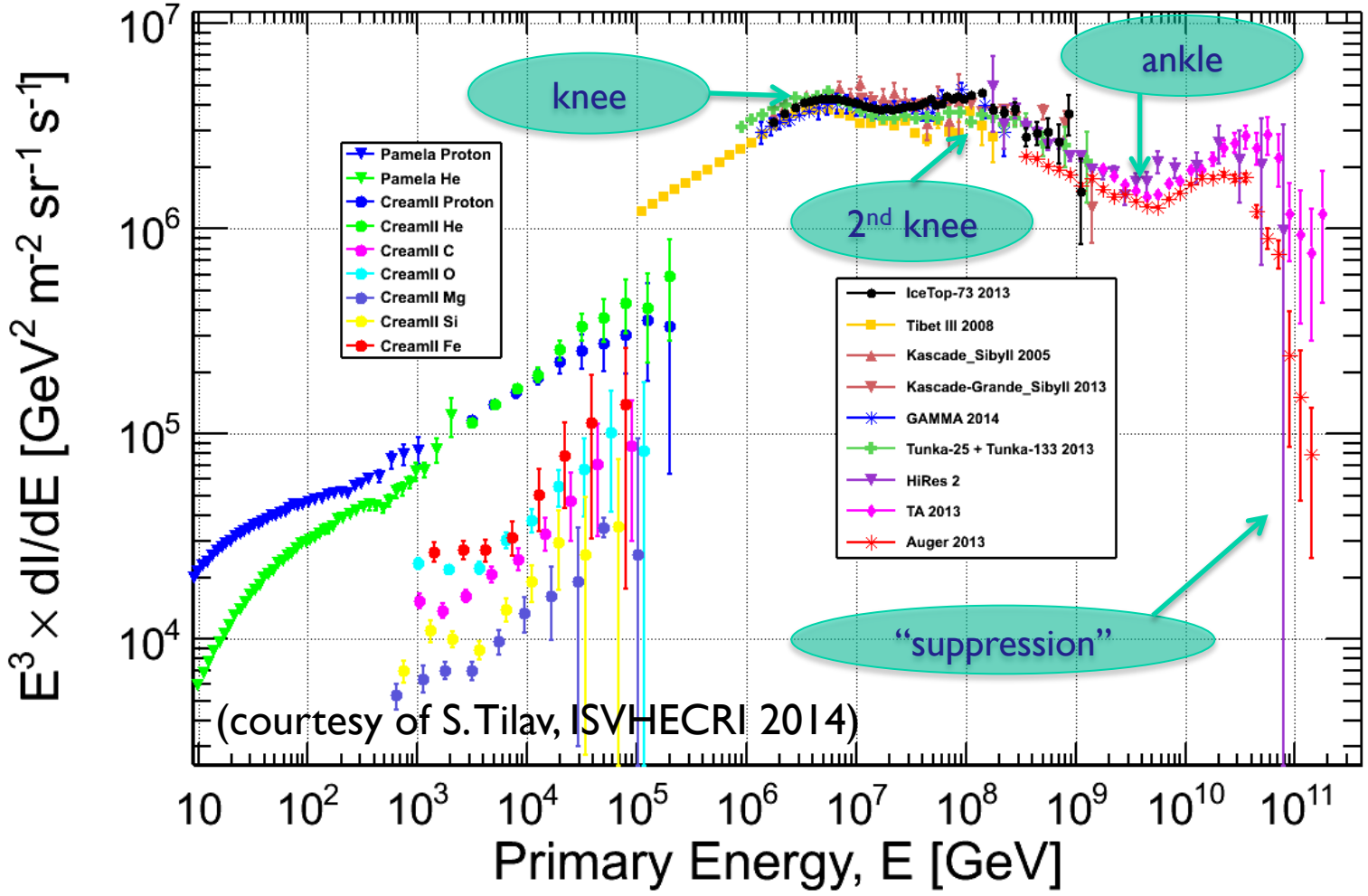
- ✓ Evidence of the most powerful astrophysical accelerators : multi-messenger astronomy
- ✓ Probe of cosmic environment through which they propagate
- ✓ Test of standard physical laws in extreme conditions
- ✓ Messengers of new physics or yet unknown particles

CR flux



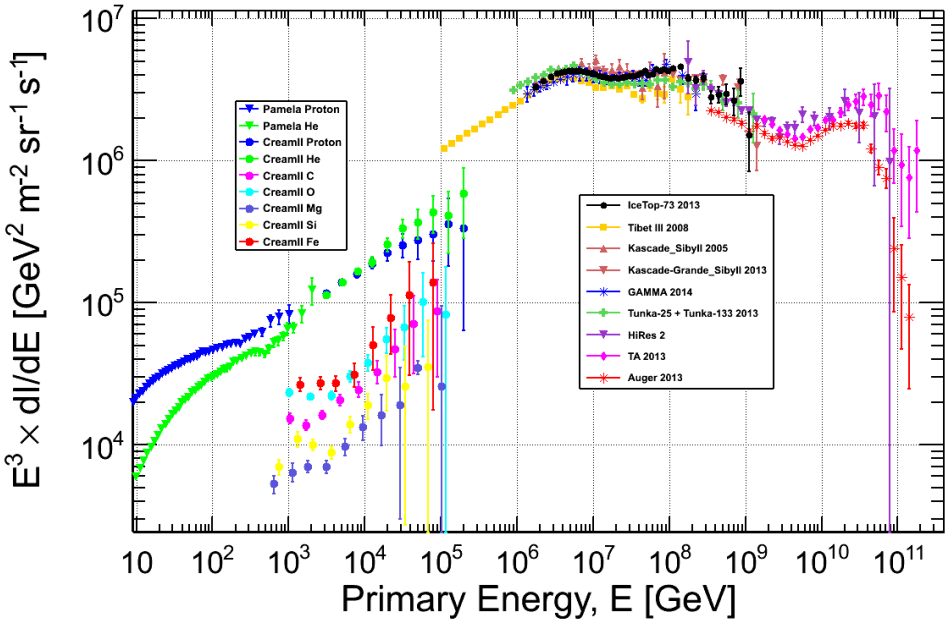
1. 11 order of magnitude in energy, 32 in flux
2. CR up to 10^{20} eV
3. Very few structures

Cosmic Ray Spectrum



Open Issues

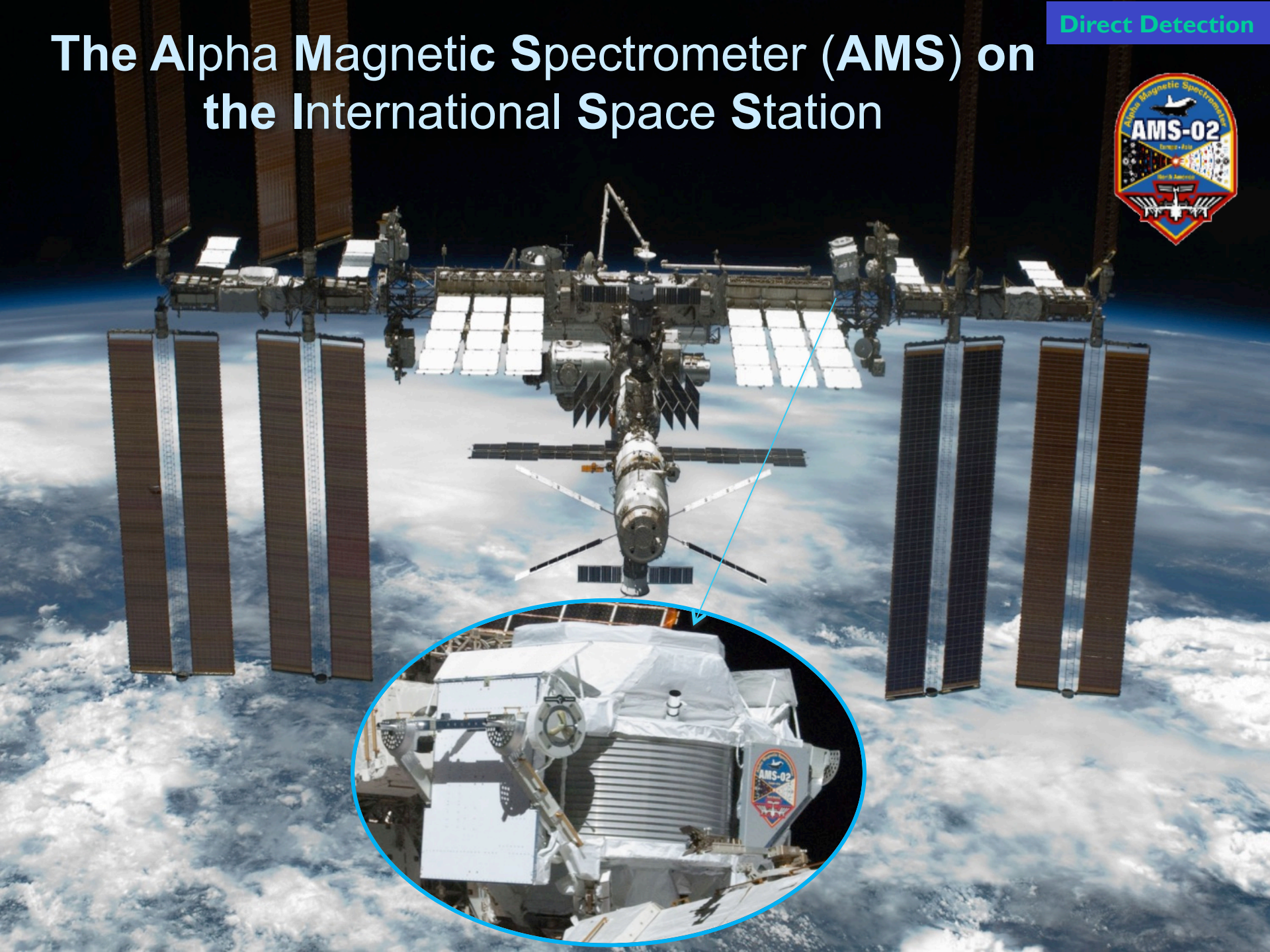
- Detailed knowledge of particle abundances below the knee: is there any hint of new physics?
- Connection between direct and indirect observations
- Nature of CR spectrum features: knee, ankle...
- Galactic to Extragalactic transitio
- The nature of the "End" of the spectrum.



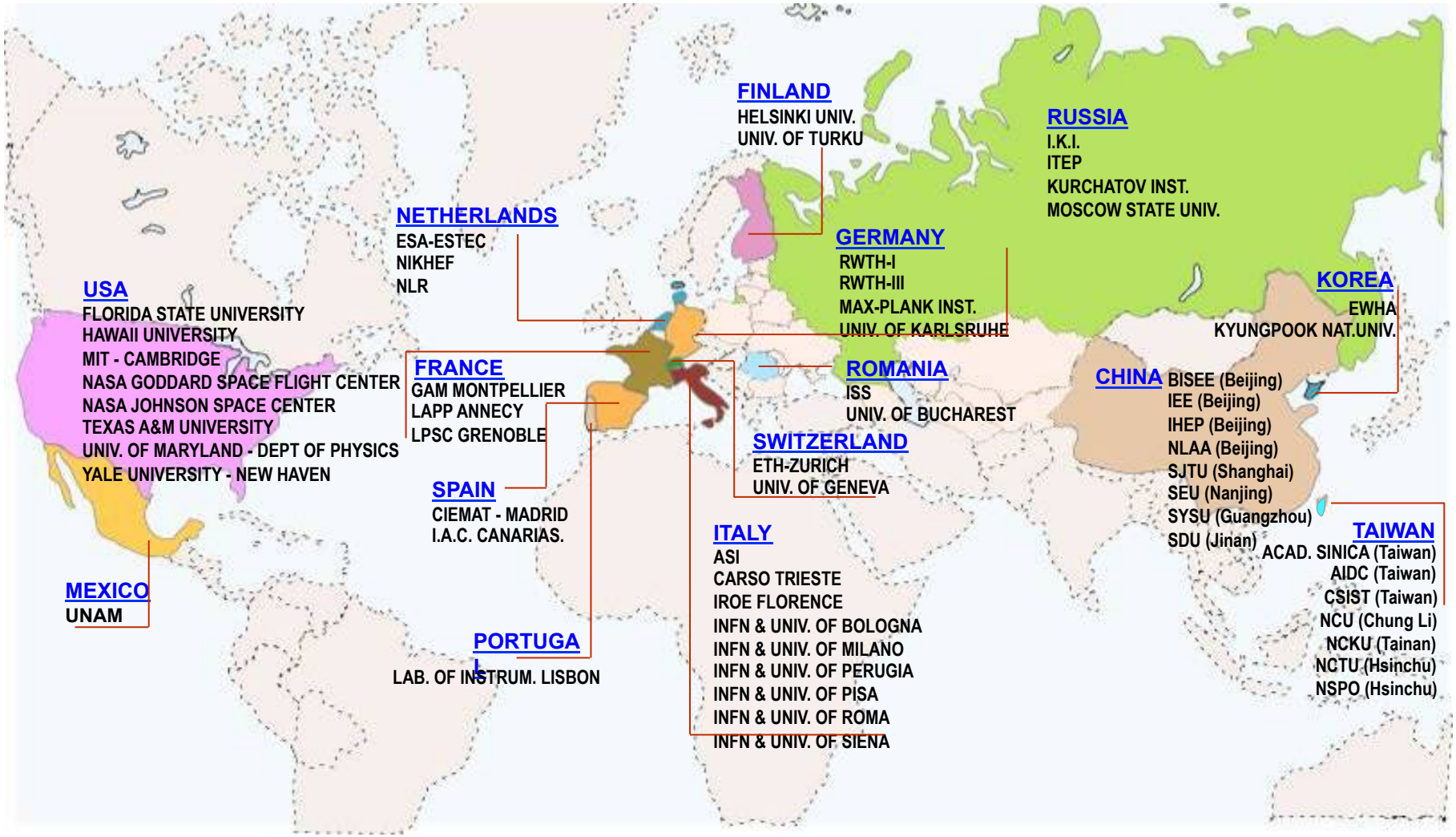
Outline

1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

The Alpha Magnetic Spectrometer (AMS) on the International Space Station

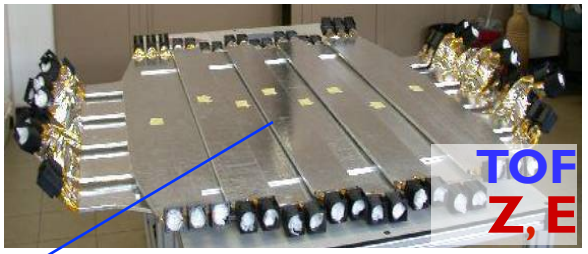


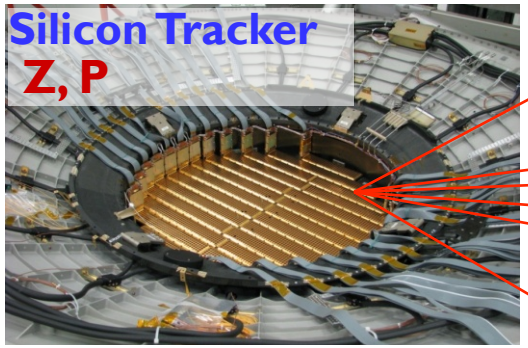
On behalf of the AMS Collaboration
16 Countries, 60 Institutes and 600 Physicists

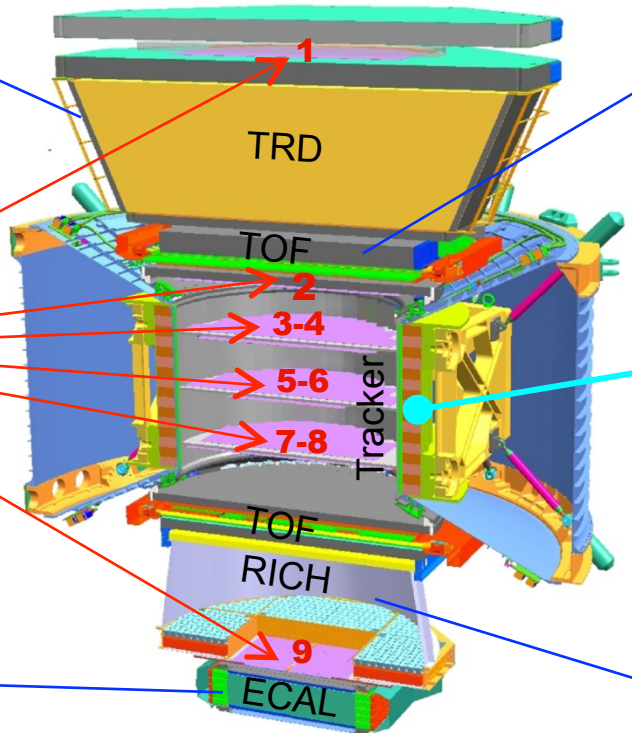


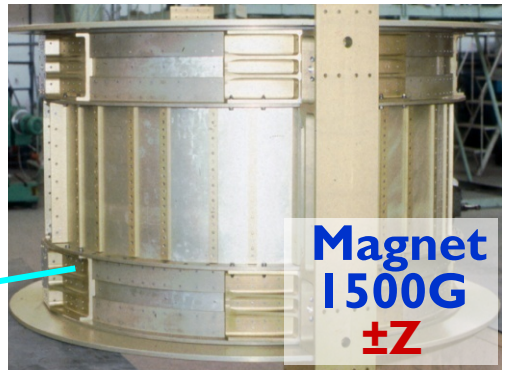
A precise, multipurpose spectrometer

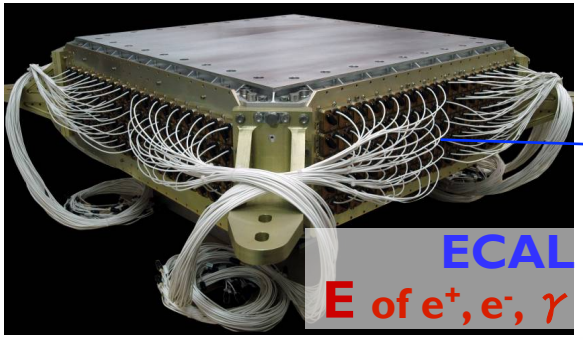
Z, P are measured independently by the Tracker, RICH, TOF and ECAL

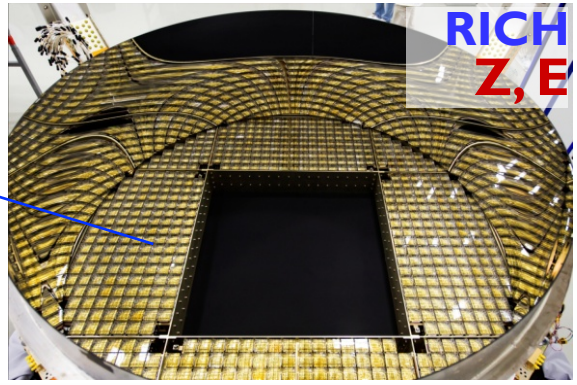




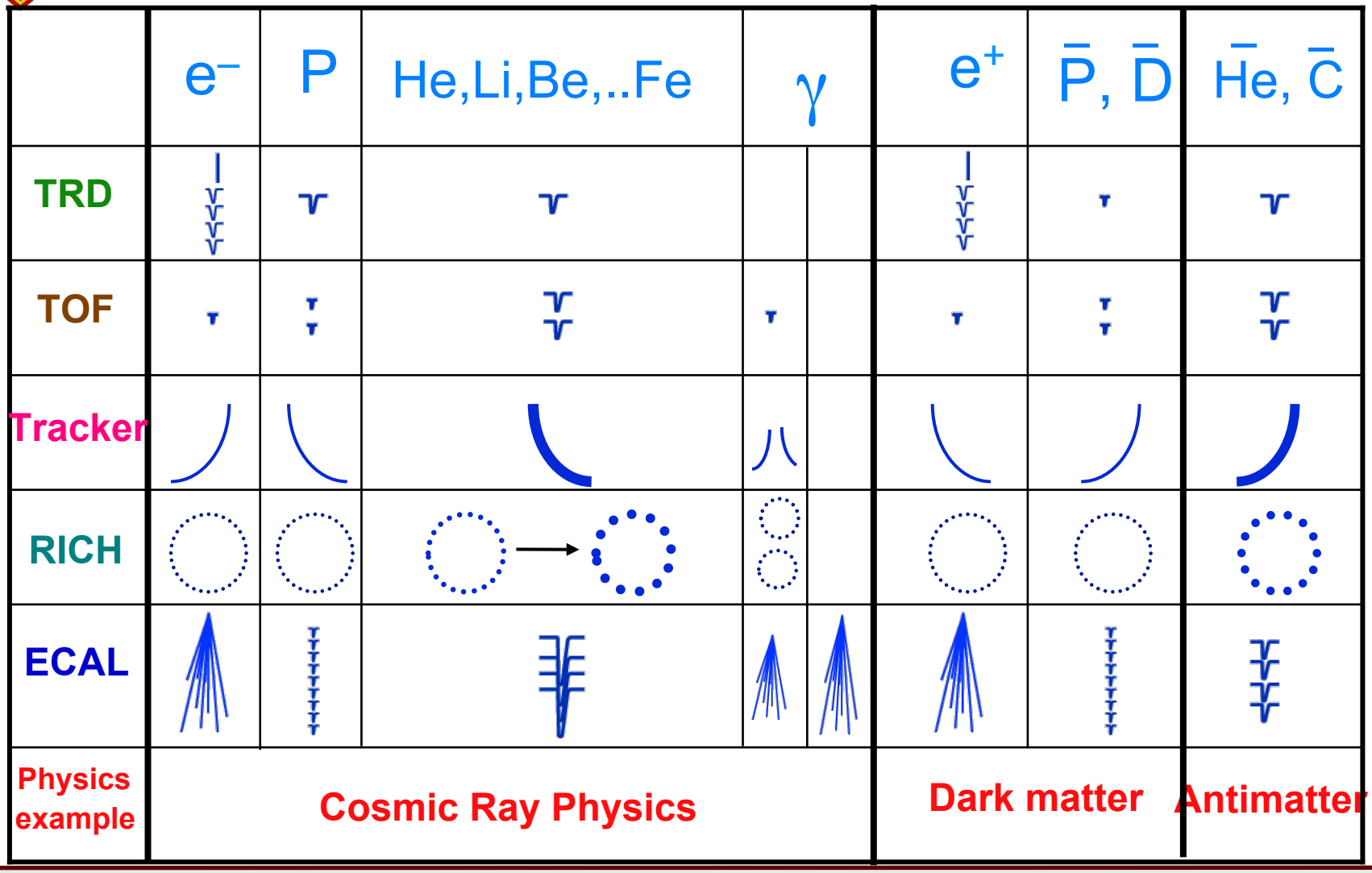




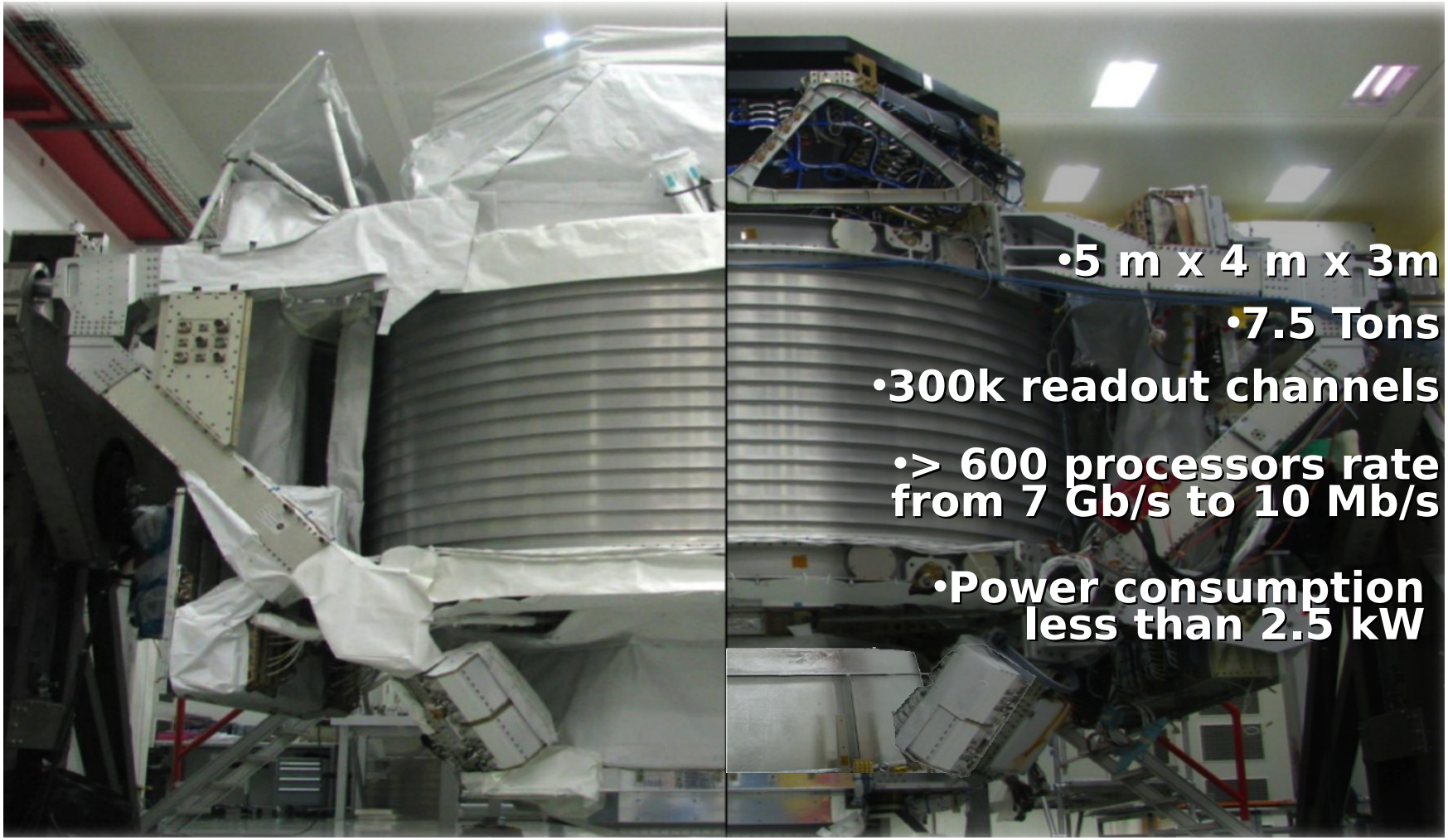




Particle Detection

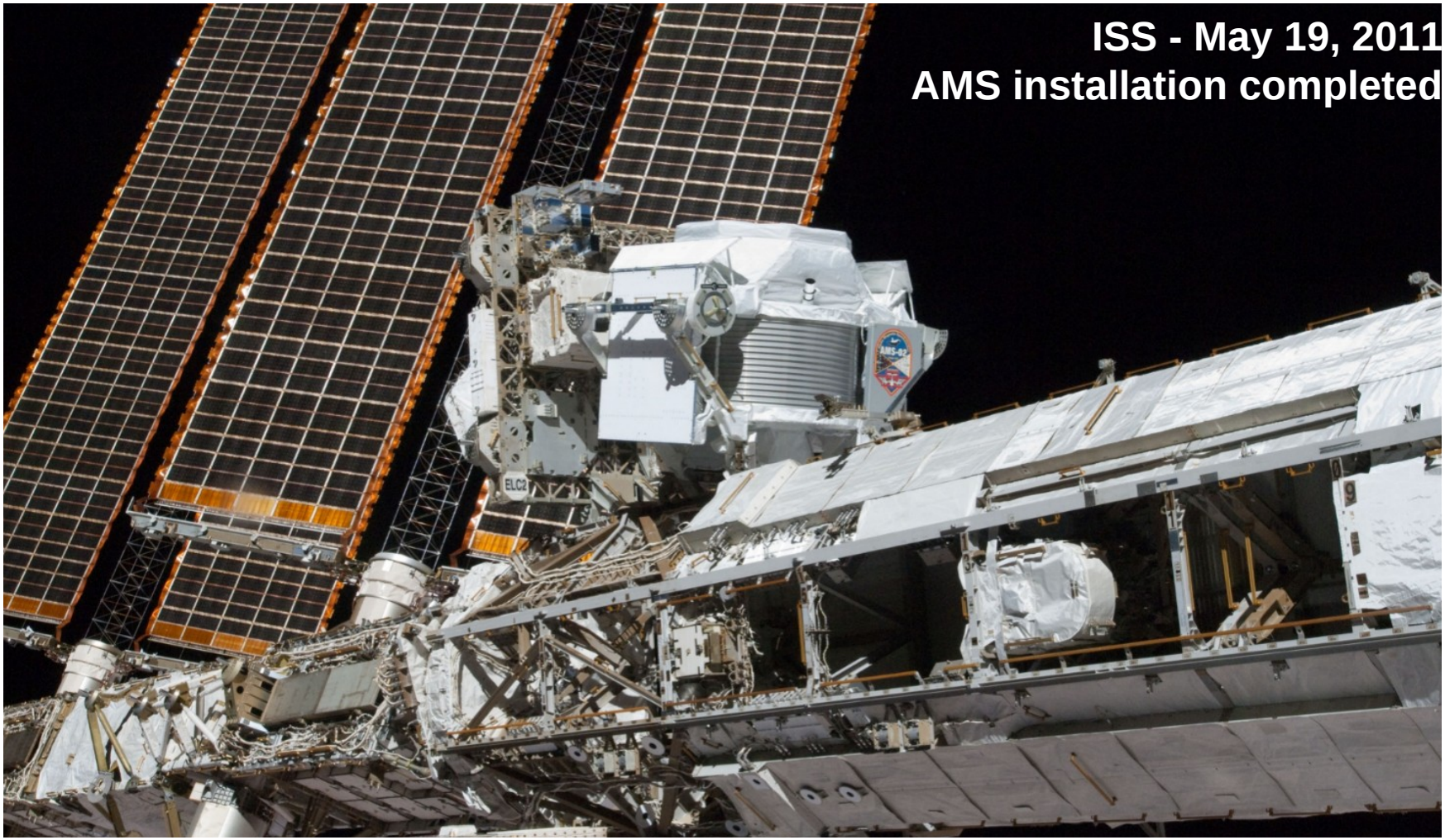


AMS02: a few numbers



May 16th 2011: AMS02 launch

May 19th 2011: on ISS



Objectives

1. Fundamental physics & Antimatter :
   - Primordial origin (signal : anti-nuclei)
   - Exotic sources (signal: positrons, anti-p, anti-D)

2. The CR composition and energy spectrum

The CR composition and energy spectrum

- Sources & acceleration : Proton and He
- Propagation in the ISM: secondaries (B/C, ...)

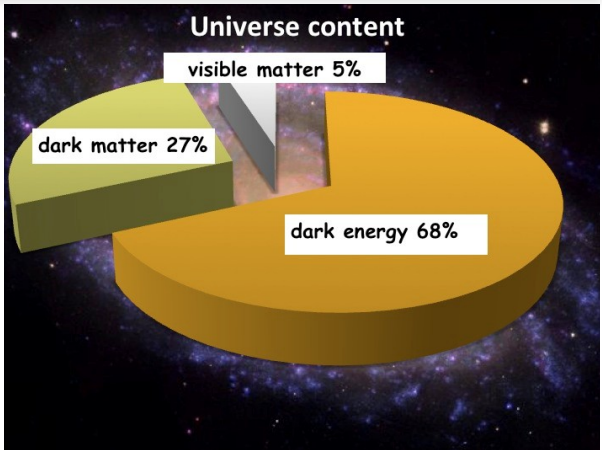
Fundamental physics & Antimatter

Primordial origin (Signal: anti-nuclei)

Dirac's Nobel speech

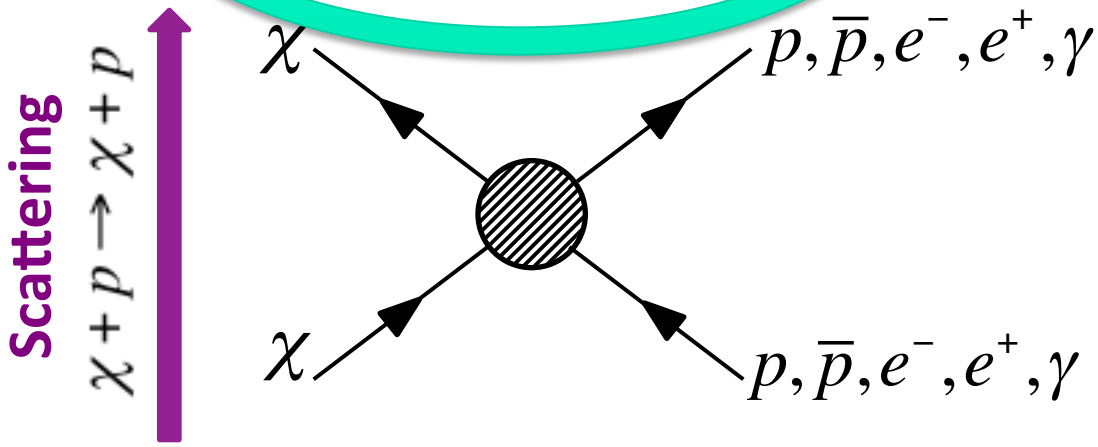
We must regard it rather as an accident that the Earth [...] contains a preponderance of negative electrons and positive protons. It is quite possible that for some stars it is the other way about

Dark Matter Quest



Annihilation

$$\chi + \chi \rightarrow p, \bar{p}, e^-, e^+, \gamma$$



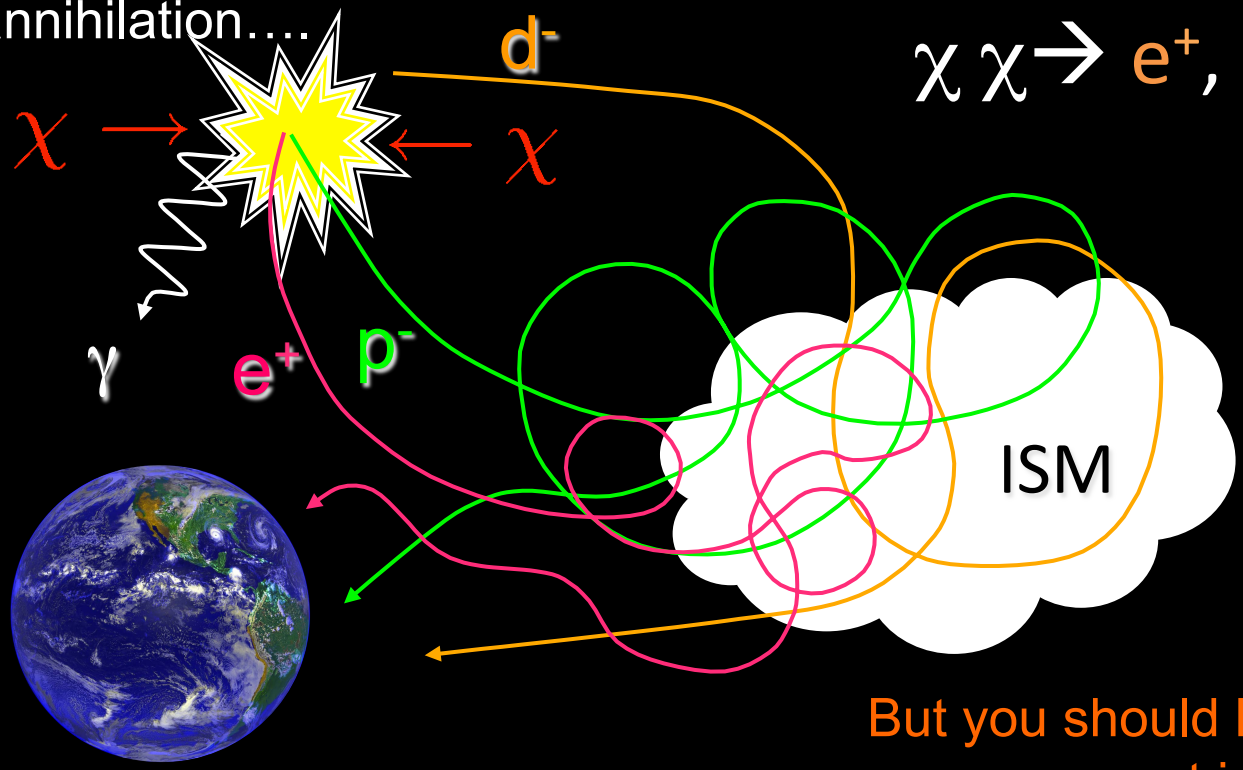
$$\chi + \chi \leftarrow p + p$$

Production

Dark Matter

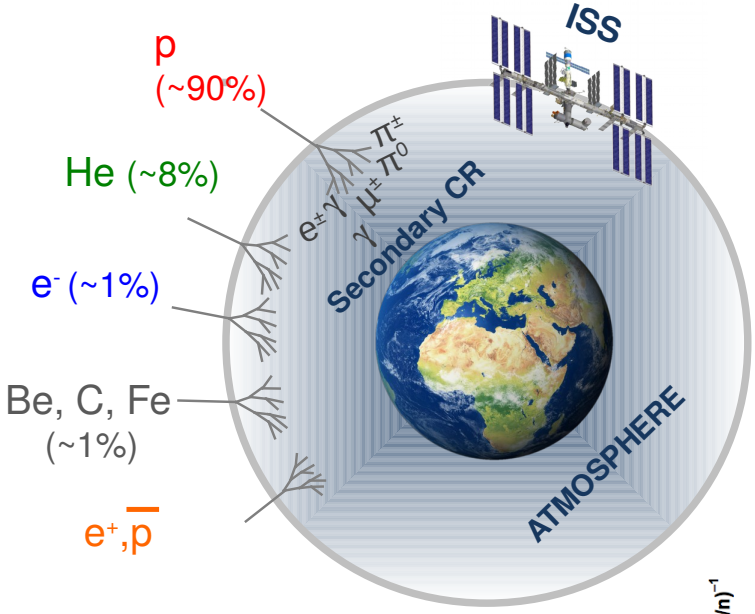
WIMP as the responsible of Dark Matter (?)

Indirect DM search → search for (rare in CR) products from their annihilation….



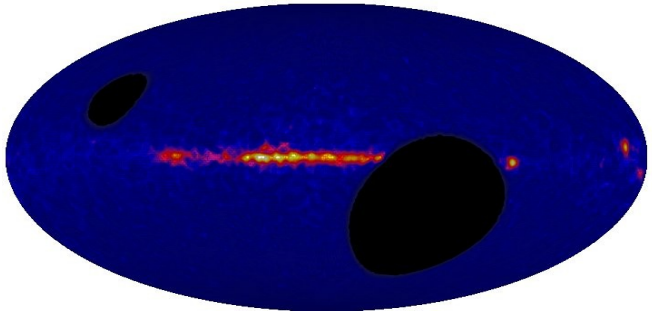
But you should know what you expect in the ISM !!

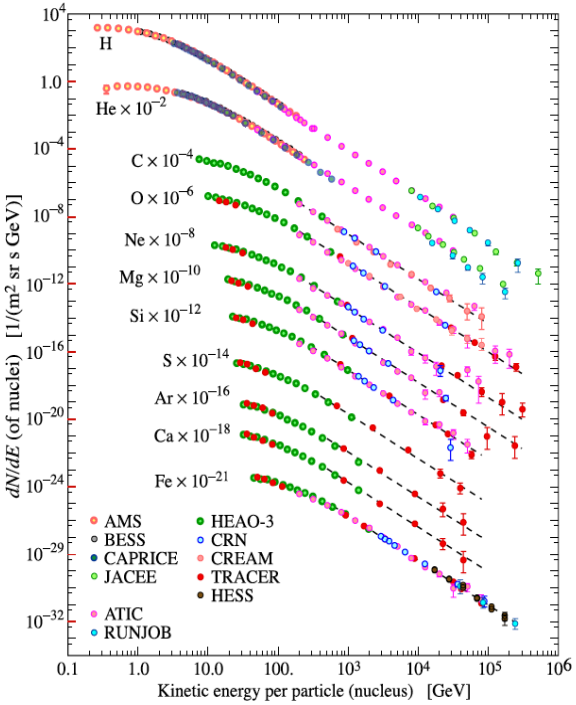
AMS measurements

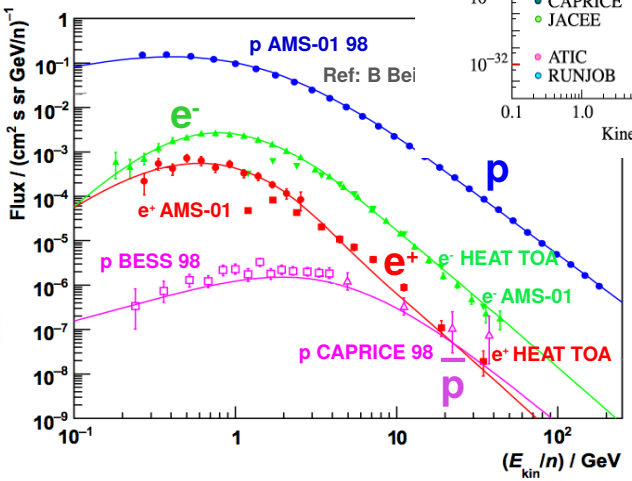


Charged cosmic rays (GV-TV)

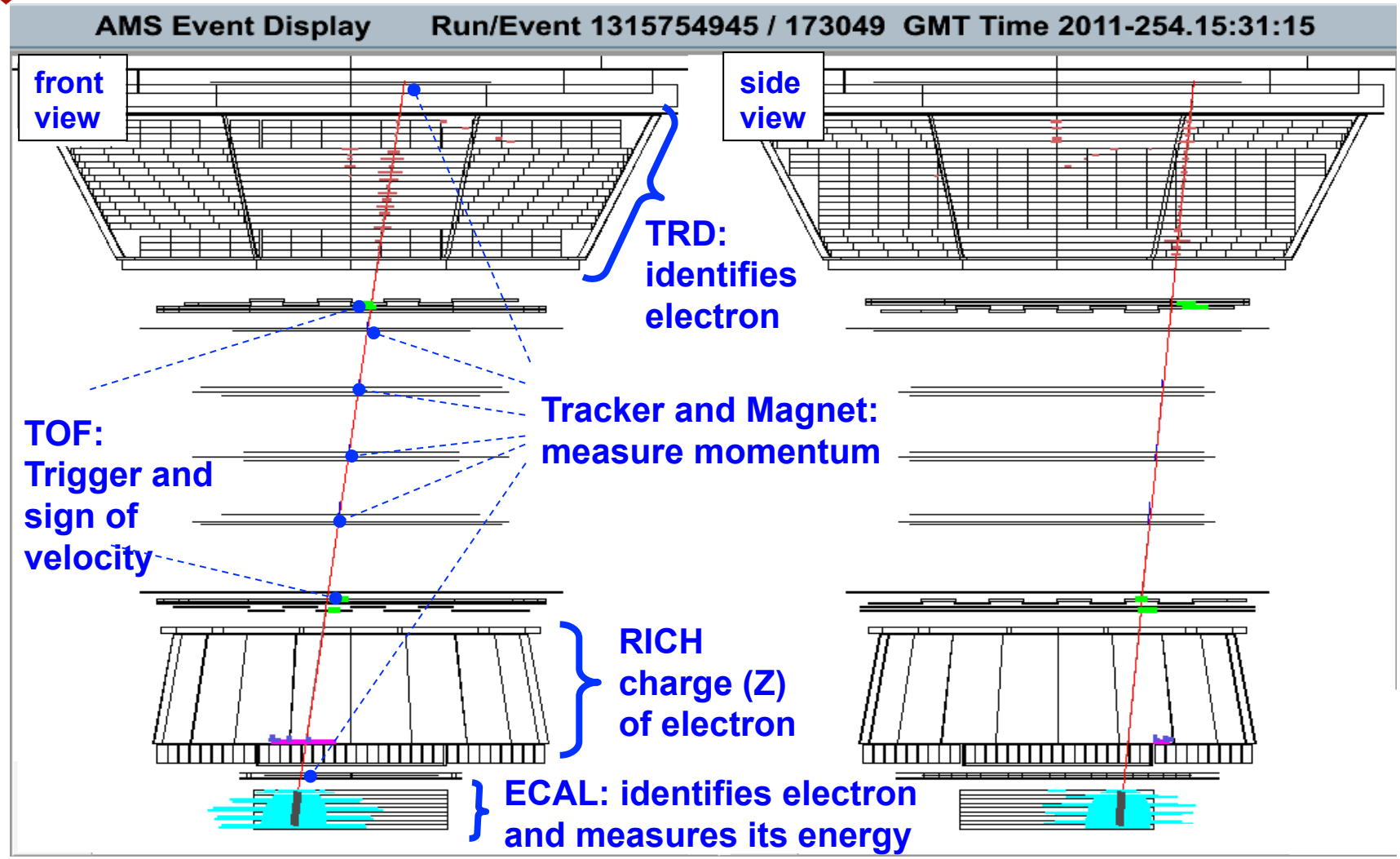
Gamma rays (E>1GeV)





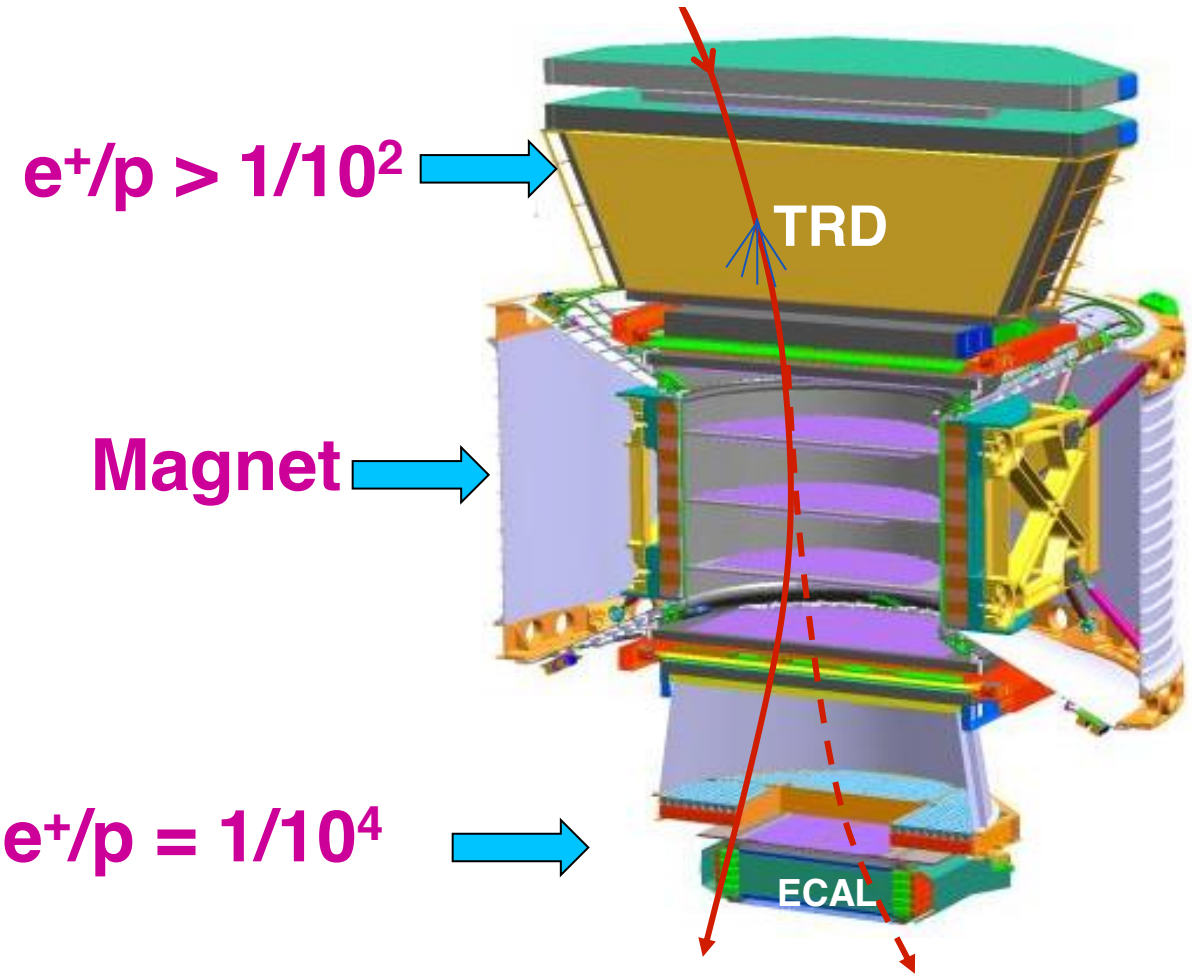


AMS data on ISS: 1.03 TeV electron



AMS Experimental Goals

$\overline{He}/He = 1/10^{10}$, $e^+/p > 1/10^6$ & Spectra to 1%



Minimal material in TRD and Tracker, so that the detector itself does not become a source of background nor of large angle scattering.

Repetitive measurements of momentum, to ensure that particles which had large angle scattering are not confused with the signal.

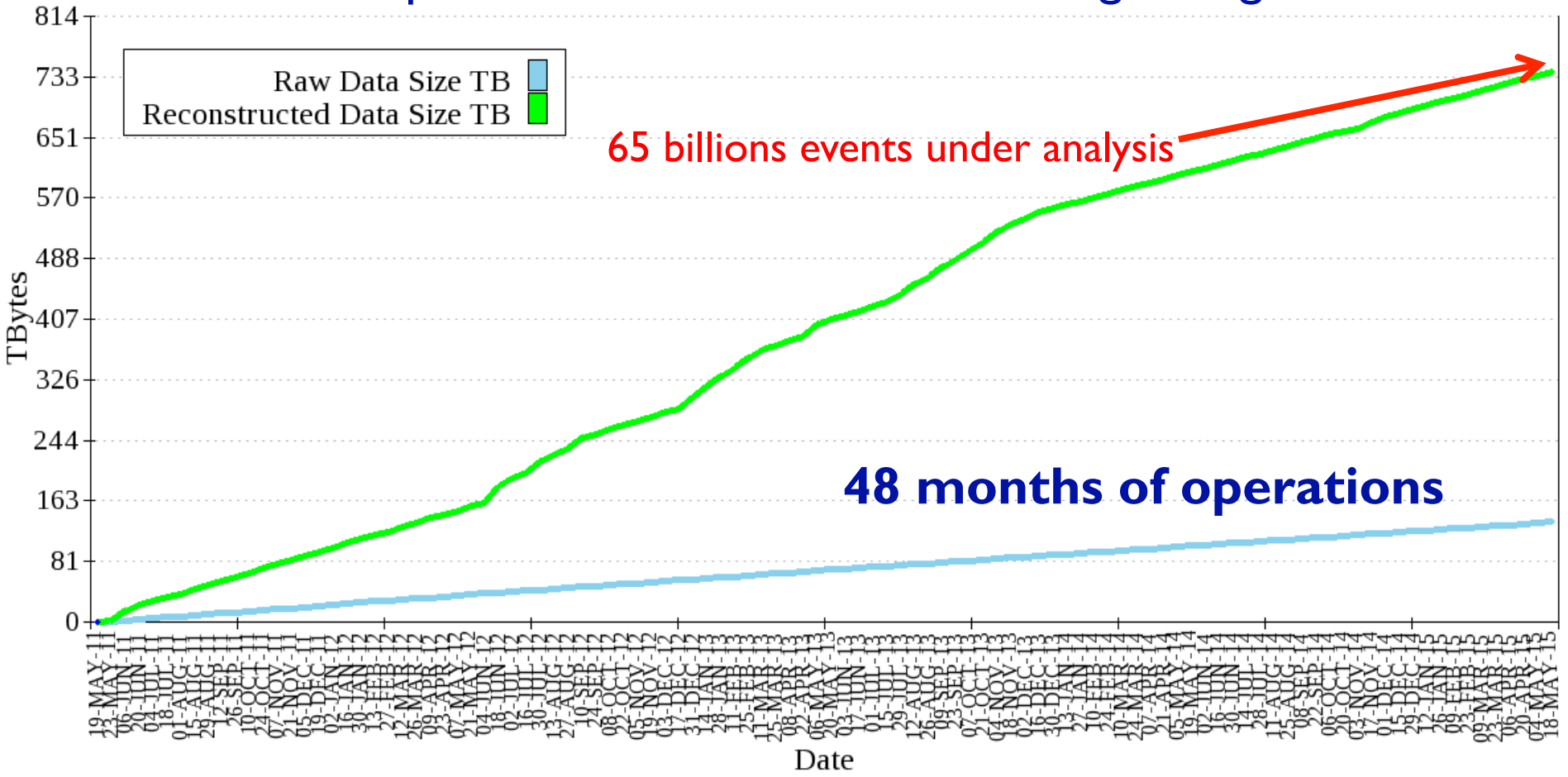
e± detectors are separated by magnetic field, so that secondary particles from TRD do not enter into ECAL.

Outline

1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

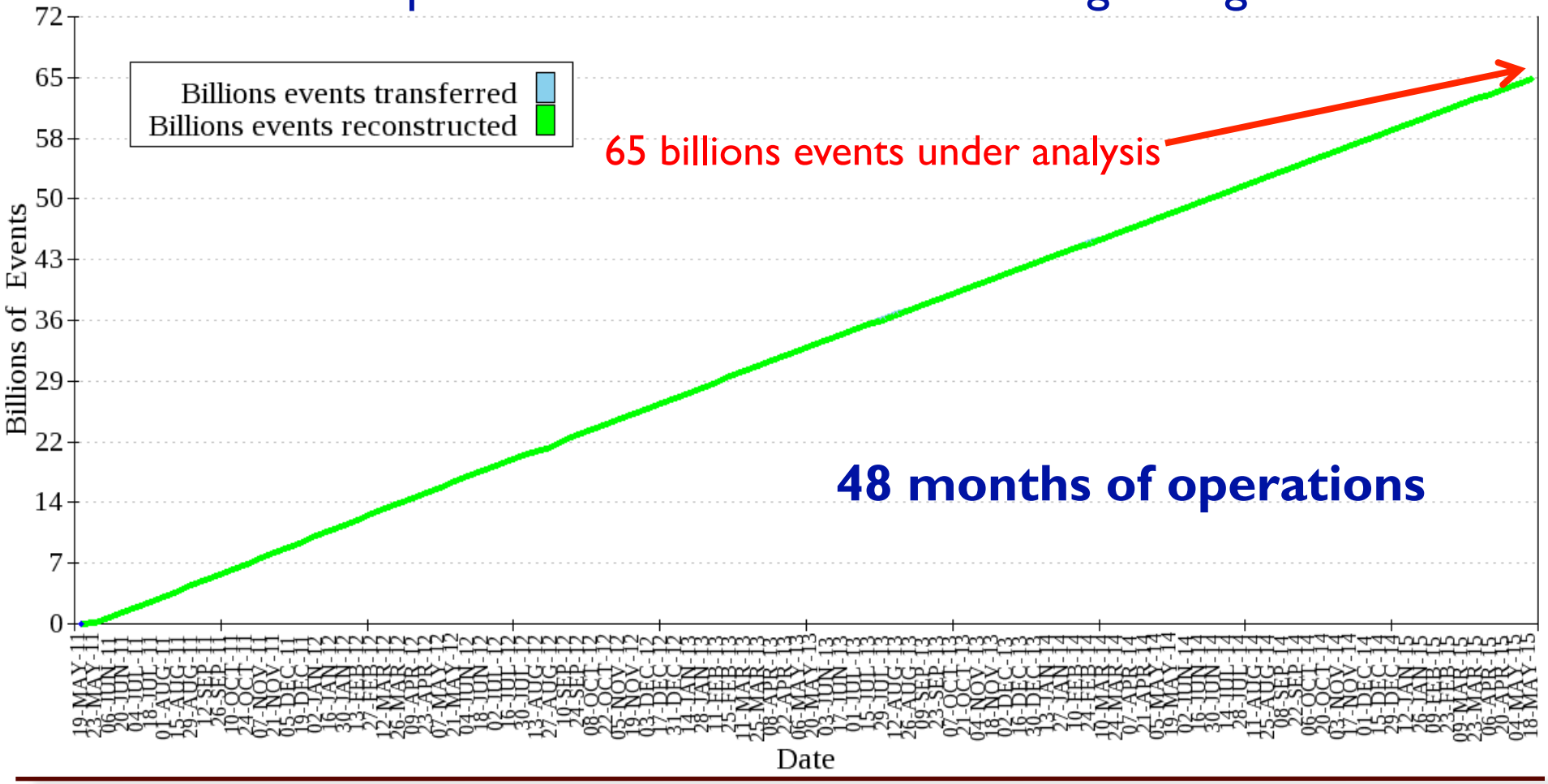
Data Collection

AMS, generates an average of 10 Mbit/s (100 GB/day).
The rate depends on the ISS location in the geomagnetic field.

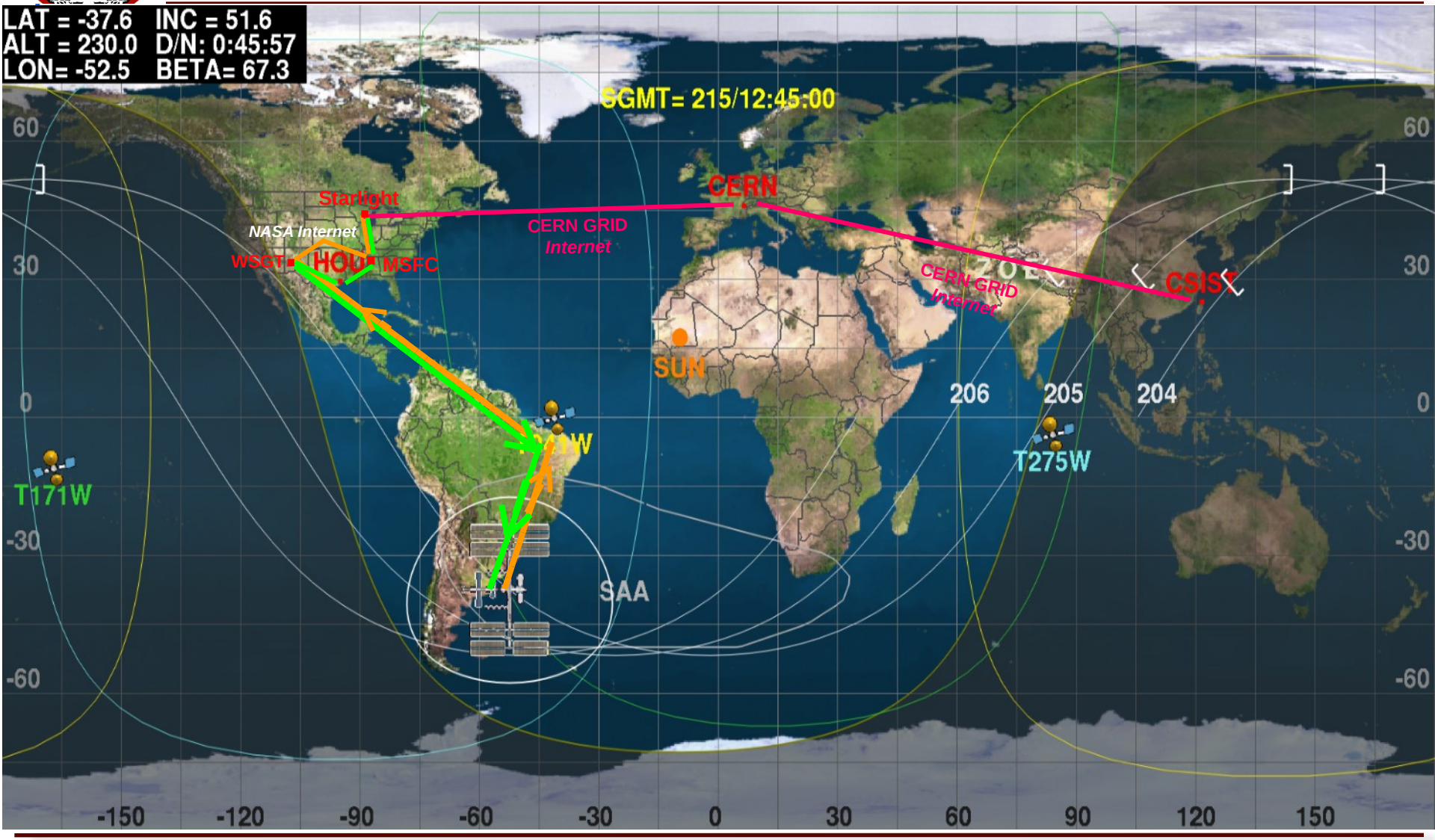


Data Collection

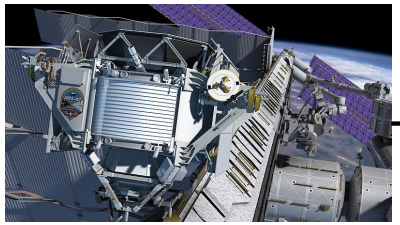
AMS, generates an average of 10 Mbit/s (100 GB/day).
The rate depends on the ISS location in the geomagnetic field.



On board of ISS @ 400 km

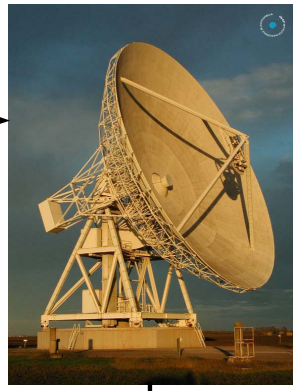


AMS data path



AMS02 on ISS

TDRSS



WSC

TAPES

POCC (CERN)

SOC (CERN)

MSFC

pcgsc50 pcgsc51

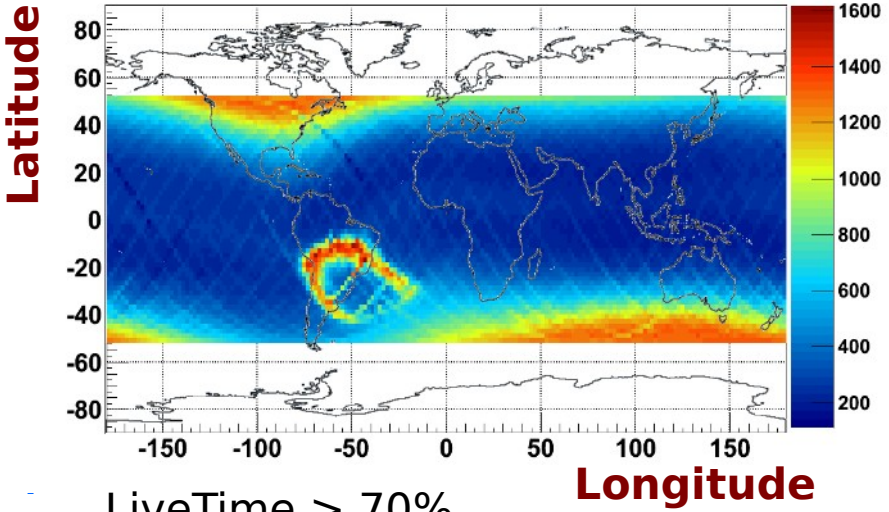
pcgsc52 pcgsc53

- TDRSS: Tracking Data Relay Satellite System
- WSC: White Sands Complex (NM)
- MSFC: Marshall Space Flight Center, Huntsville (AL)
- POCC: Payload Operations Control Center
- SOC: Science Operation Center
- Partial links availability
- Data loss

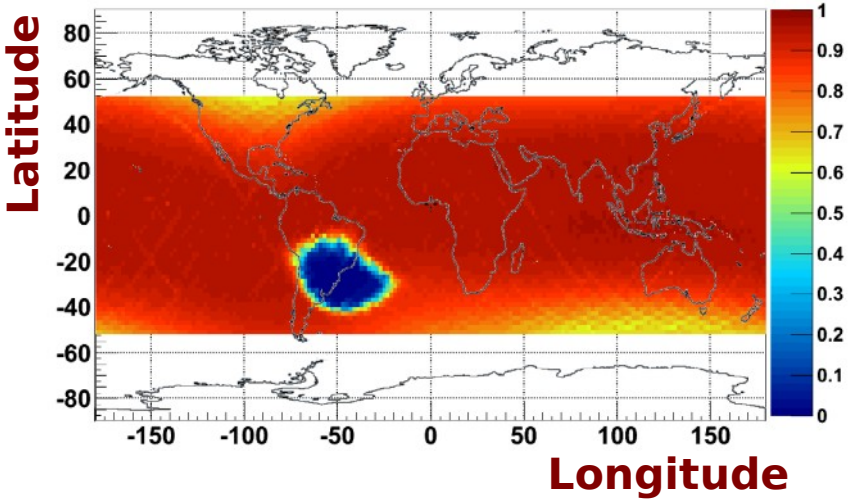
Two Comunication lines: **Ku** (12-18 GHz)band and **S** (1.55-5.2 GHz) band

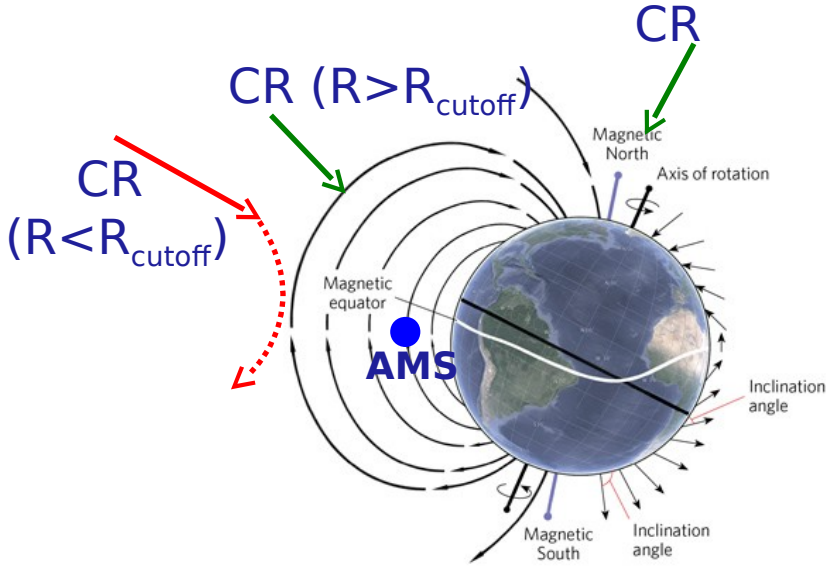
Data taking 24h 7/7 days

<Acquisition rate> ≈ 500 Hz



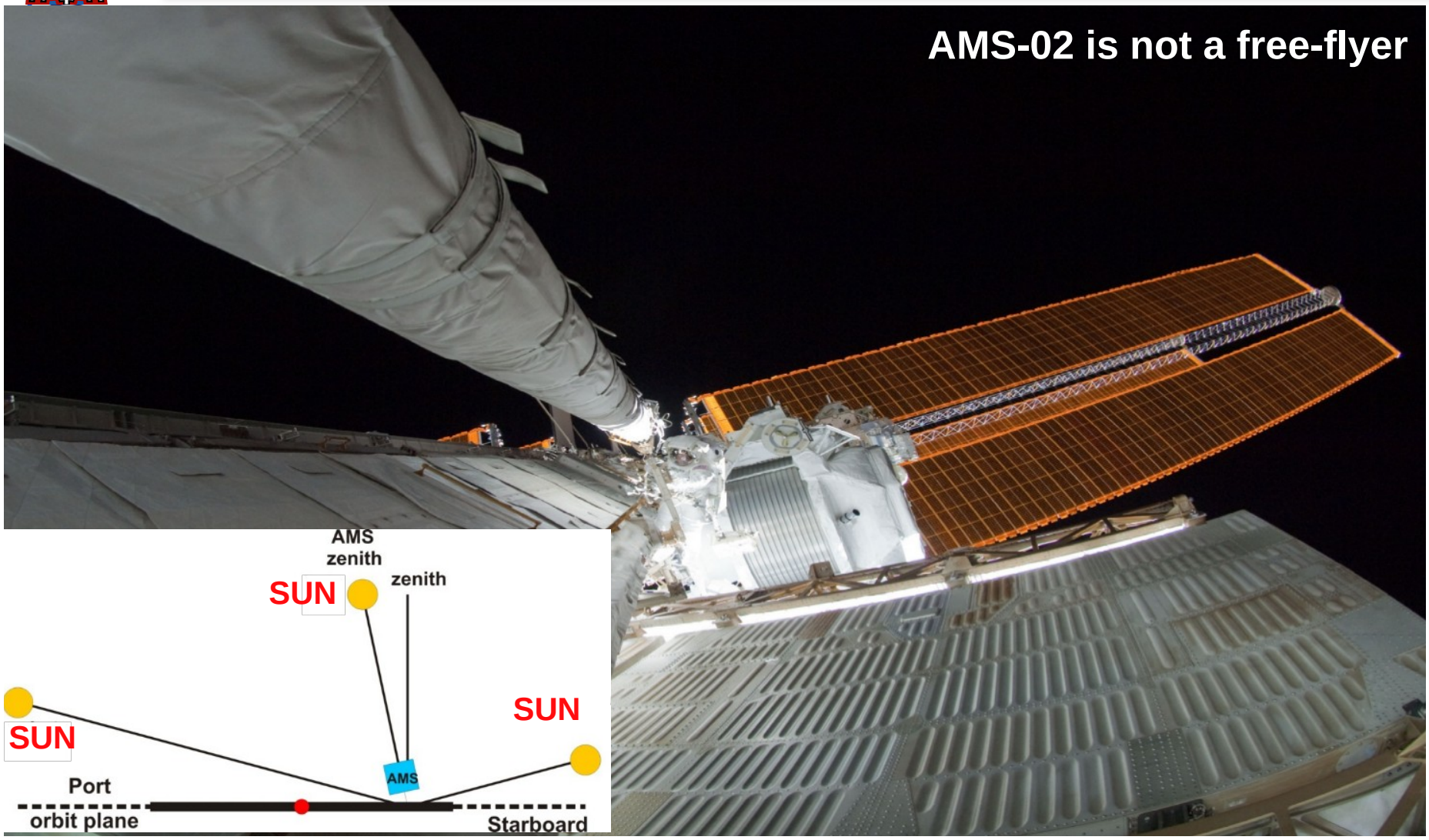
LiveTime > 70%





Payload Operation Control Center (POCC) @ CERN

The thermal environment

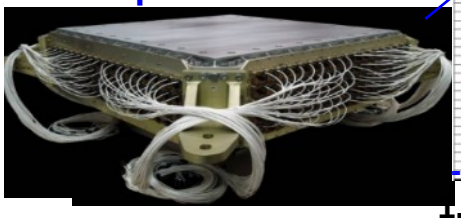


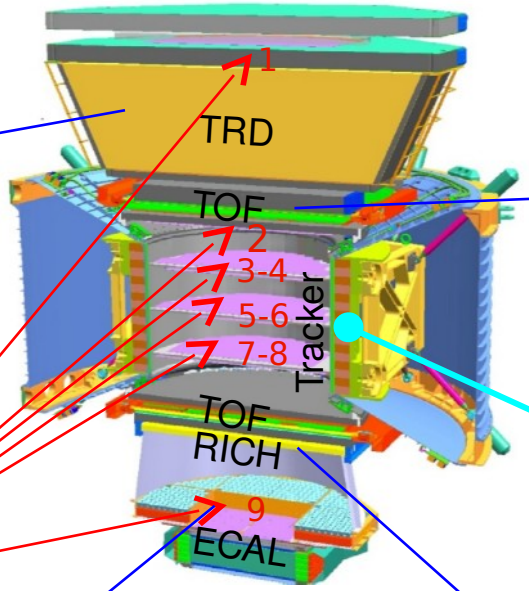
T/P sensors + heaters

TRD
24 Heaters
8 Pressure Sensors
482 Temperature Sensors

Silicon Tracker
4 Pressure Sensors
32 Heaters
142 Temperature Sensors

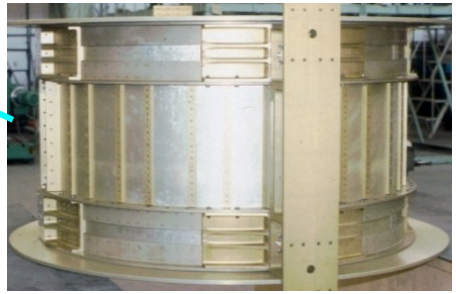
ECAL
80 Temperature Sensors



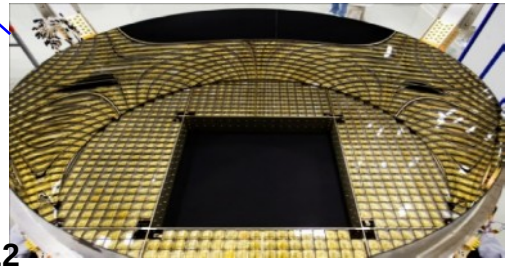


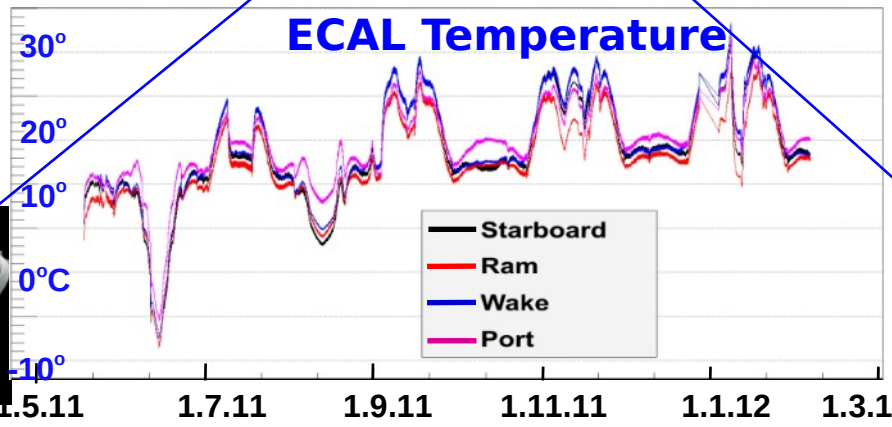
TOF & ACC
64 Temperature Sensors

Magnet
68 Temperature Sensors



RICH
96 Temperature Sensors





On line Monitoring

Outline

1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

AMS Data structure

- AMS has to be compliant to all of the standard communication protocols used by NASA to communicate with ISS, and its data have to be transmitted through the NASA communication network.

- data are organized in a custom format called AMSBlocks that can assume many forms, such as special blocks for science data, or envelope blocks containing a set of sub blocks

AMS Data structure

- AMSBlocks are divided into smaller pieces of fixed size: the frames. Format is defined by the Consultative Committee for Space Data System (CCSDS)

- AMS detectors produce AMSBlocks that are split in pieces, one minute frames. Frames are transferred via NASA communication system and copied locally on ISS, on a dedicated laptop.

- At CERN, frames joined to rebuild up the AMSBlocks.

AMS Data structure

- Frame: uniquely identified by the triplet *(APID, Time, SeqNo)*
- At merging time, completeness of data stream is verified and a play back from AMS laptop of missing or corrupted frames is performed
- AMSBlocks are then merged to build up runs (raw AMS data). Each run lasts almost 23 minutes, one quarter of the ISS orbit. Finally, raw data are reconstructed and a ROOT format is produced

AMS Data structure

1. Two independent data productions on board
   - House Keeping Data (HK), ~ 20 Kb/s
   - Science Data (SCI) ~ 6-24 Mb/s
2. Two external interfaces
   - HRDL (High Rate Data Link)
   - LRDL (Low Rate Data Link)
3. HK Data always goes on both interfaces
4. SCI data has two possibilities
   - go out immediately after being produced
   - enqueued into the a memory buffer

Outline

1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

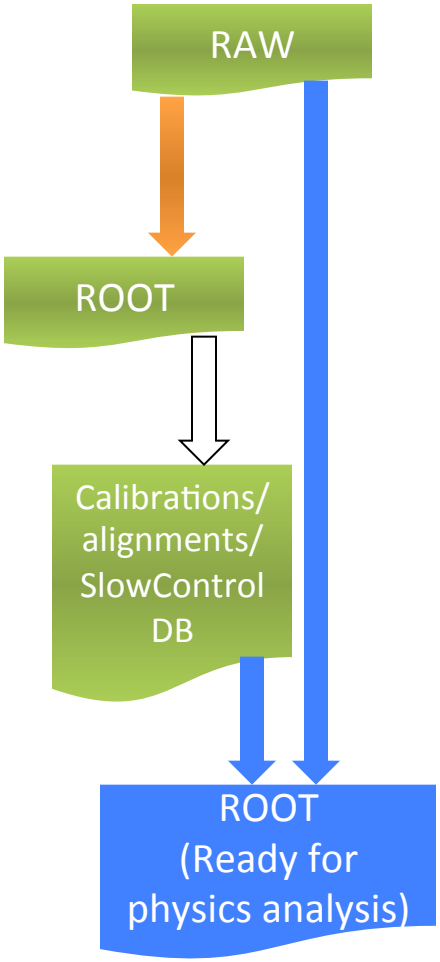
Online Reconstruction

- Science stream is decoded on a local cluster and processed to produce
  - Calibration Data for the tracker silicon detector
  - Simplified reconstruction data stream for each one minute frame
- Local POCC production is used to monitor sub-detectors behavior and to verify their performances as a function of time
- One-minute and Calibration production are usually available at most 5 minutes later than the data arrived

Data Reconstruction

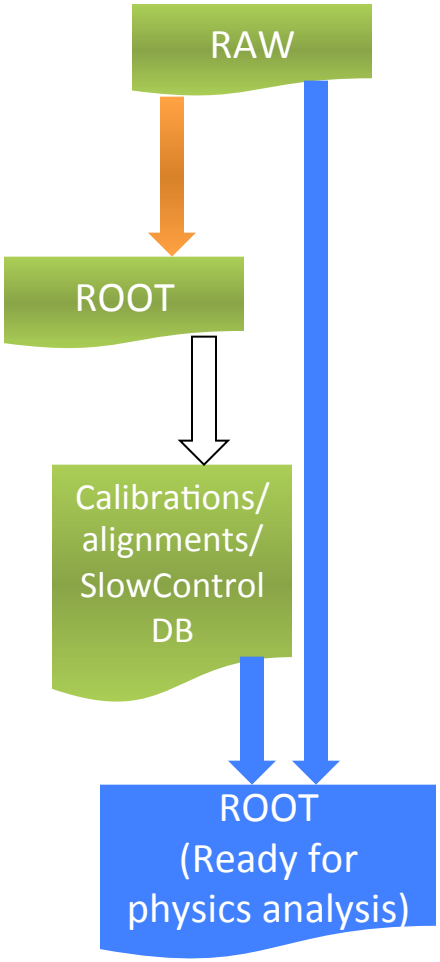
Data reconstruction pipeline is mainly composed by two logical step

1. First Production (std)
   - ✓ Runs 7x24 on freshly arrived data
   - ✓ Initial data validation and indexing
   - ✓ Produces Data Summary Files and Event Tags (ROOT) for fast events selection
   - ✓ Usually be available within 2 hours after flight data arriving
   - ✓ Used to produce various calibrations for the second production as well as quick performance evaluation



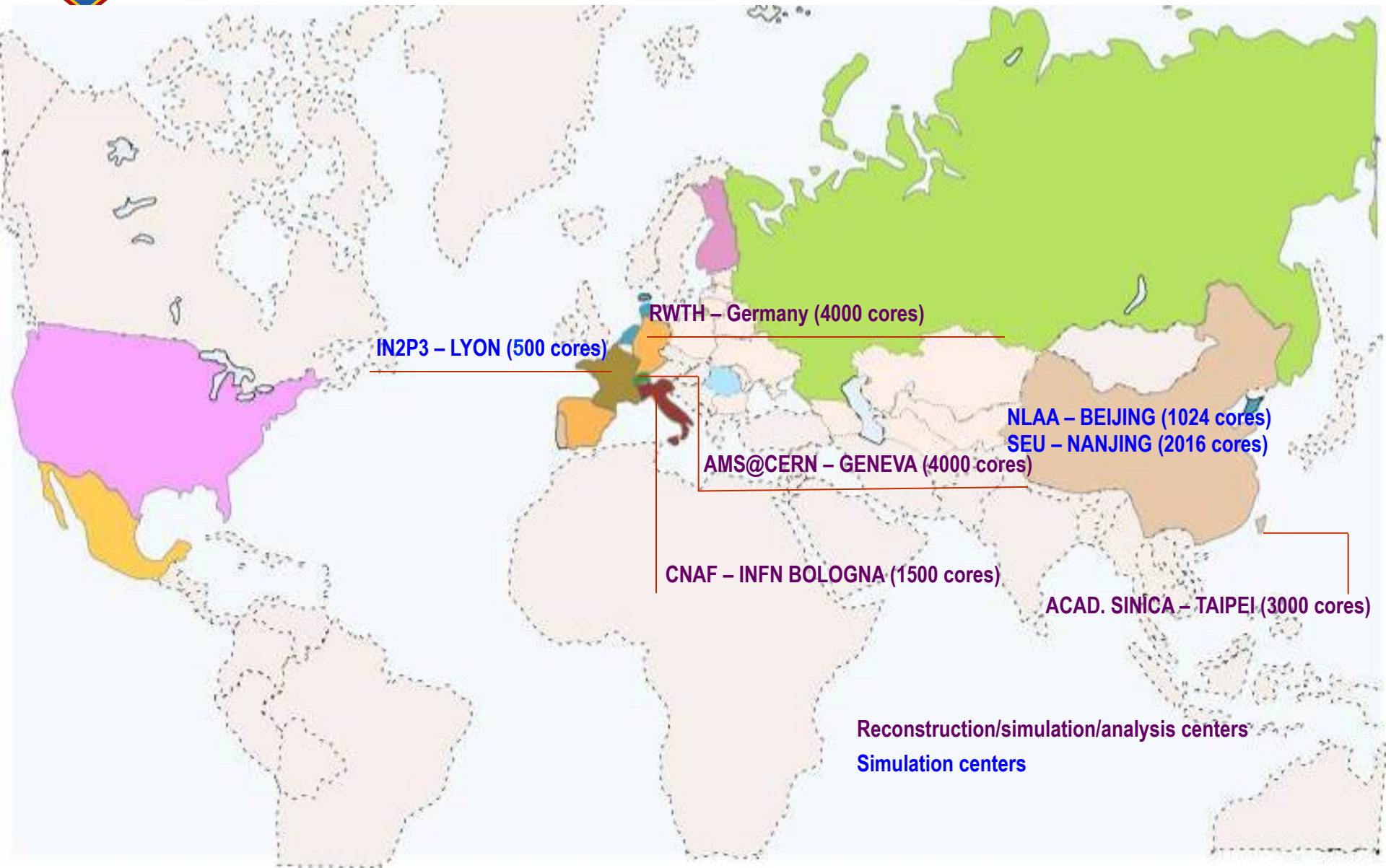
2. Second Production

- ✓ Every 3-6 months incremental
- ✓ Full reconstruction in case of software major update
- ✓ To use all the available calibrations, alignments, ancillary data from ISS, and slow control data to produce physics analysis ready set of reconstructed data



The First Production is processed on a local AMS farm at CERN of about 200 cores, whereas Monte Carlo productions, ISS data reprocessing and user data analysis are supported by a network of computing centers

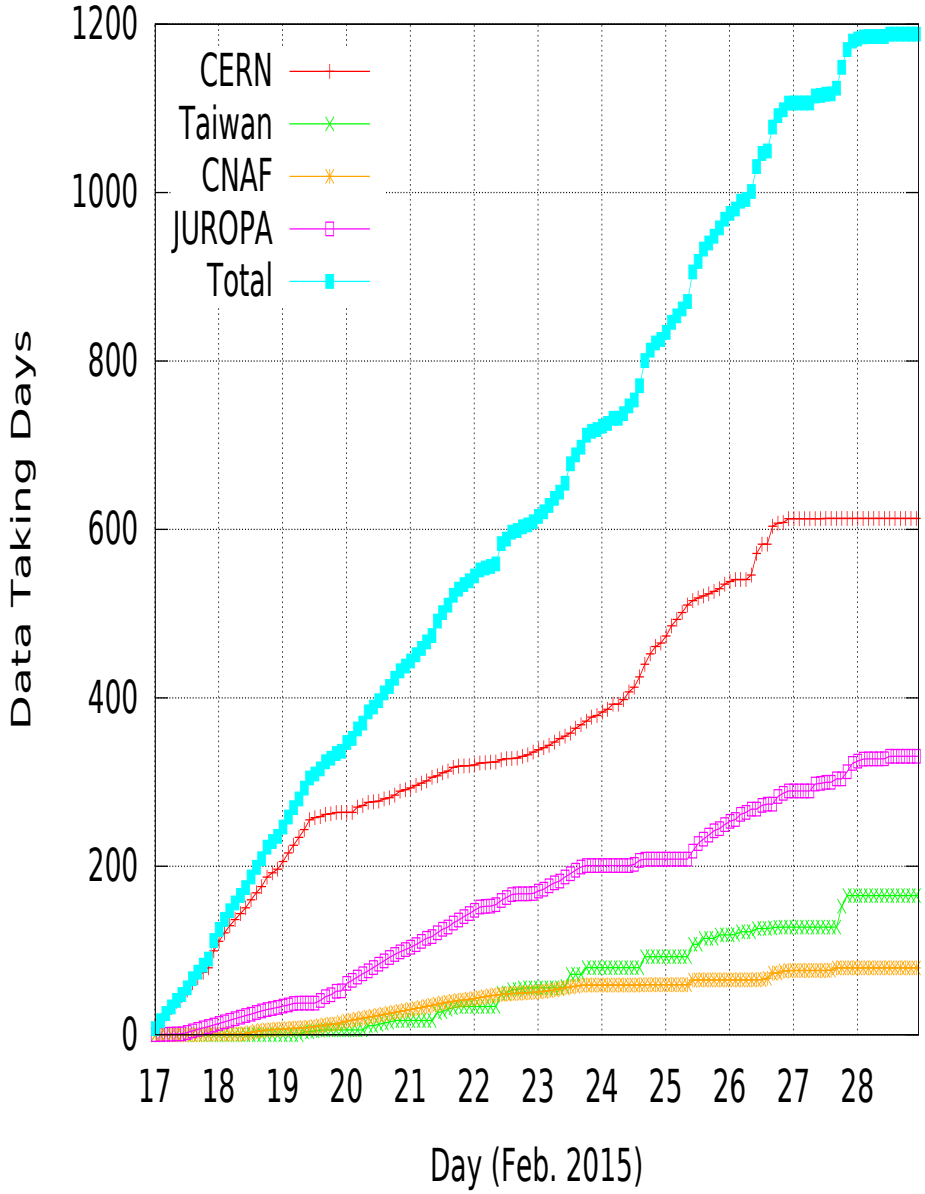
AMS-02 Major Computing Contributors



Light-weight production platform

1. Fully-automated production cycle
   - ✓ Job acquiring, submission, monitoring, validation, transferring, and (optional) scratching
2. Easy to deploy
   - ✓ based on Perl/Python/sqlite3
3. Customizable
   - ✓ Batch system, storage, transferring, etc.
4. Running at:
   - ✓ JUROPA, CNAF, IN2P3, NLAA

Latest Full Data Reconstruction



- ✓ Second half of Feb. 2015
- ✓ Performed at CERN, JUROPA, CNAF, ACAD. SINICA, and all provided exclusive resources
- ✓ 40 months' flight data
- ✓ Reconstruction completed essentially in two weeks
- ✓ 100 times data collection speed
- ✓ 10^{-5} failing rate

Reconstruction Parallelization

- Motivations
  - ✓ To shorten the (elapsed) time of data reconstruction
  - ✓ To increase productivity by using multi-core/ hyper-thread enabled processors
  - ✓ To reduce memory usage per core
  - ✓ To ease the production jobs management
- Tool: OPENMP
  - ✓ No explicit thread programming
  - ✓ Nearly no code change except few pragmas

Simulation Code Parallelization

- Motivations
  - ✓ To decrease required memory amount per core
  - ✓ To shorten elapsed running time so as to allow faster Monte-Carlo parameter/software tuning
- Minimal software change:
  Original Geant4.10.1 MT model + OPENMP (ICC 15)
  - ✓ Thread ID, thread number, and core number are taken from Geant4.10.1, instead of from OPENMP
  - ✓ Barriers from OPENMP not working (custom implementation)
  - ✓ Other OPENMP pragmas work (minimal changes to AMS software)
  - ✓ No changes to original G4.10.1

Outline

1. Introduction to Cosmic Ray physics
2. AMS02: Detectors and Physics Goals
3. The Experimental Challange:
   - Detector Monitoring
   - Data structure and data path
   - Data production
   - Data access
4. Summary

Data Access

- Reconstructed data are distributed from the SOC (Science Operation Center) to Regional Sites that will act as storage of AMS for the user analysis.
- At CERN data are copied to CASTOR and are also available on EOS.
- Calibration and time dependent detector information are available trough the afs and cvmfs.

INFN contribution

- CNAF is part of the AMS computing center network
  - ✓ ~ 8000 HS06 and ~1.5 PB of disk space
- Additional 10000 HS06 have been pledged during the last data re-processing
- Two local queues available for INFN AMS users:
  - ✓ ams, CPU limit=3300 minutes, maximum number of running job=600;
  - ✓ short, with an higher priority but a CPU limit=360 minutes and a max number of running job=100
- ams prod queue, CPU limit=5760 minutes and 2000 jobs limit, is available only for the production team

Data Access for the INFN groups

- A Data Transfer software has been setup and is maintained to efficiently transfer data from (to) CERN to (from) CNAF
  - ✓ no interference with SOC activities
  - ✓ it keeps track of what has been moved and if successfully or not.
- The main core of the DT is a Multi-threaded finite state automaton (written in Python) and the state transition jobs are written in Perl.
- It uses a database (Mysql/Oracle) for book-keeping and it relies on GRID's file transfer protocols

Remote Data Access

- Most users are locally running at CNAF.
- A local farm is available in ASDC for end point analysis, with relevant CPU resources.
- It has implemented an integrated solution, which enables transparent and efficient access to on-line and near-line data through high latency networks.
- It is based on the use of the General Parallel File System (GPFS) and of the Tivoli Storage Manager (TSM).

Remote Data Access

- A single, geographically-distributed namespace, characterized by automated data flow management between different locations has been defined (thanks to the Active File Management, AFM, of GPFS)

- A database (based on ROOT TTree objects) with tags of events passing certain preselection requirements has been locally created in ASDC.

Remote Data Access

- Each data processing job at ASDC queries the preselection database to look for the tags of interesting events, to access them (and only them) from a remote file.
- AFM Prefetch Threshold has been tuned to manage 10 GB files (average size of AMS run file) accessed randomly.
- The configuration allows to process the same file remotely paying only a fraction of 15% in execution time

Summary

- CR are an unique window over new physics and a highway to the cosmology: DM and anti-matter searches

- AMS is the Cosmic Rays observatory of the next decade, 65 billion of events collected in 4 years of operation on the International Space Station.

- AMS experiment has a need of computing resources similar to LHC experiment: in terms of HS06 and disk space for data processing and Monte Carlo production

Summary(2)

- AMS detector lives in hostile environment and need to be continuously monitored.
- AMS data are transmitted from ISS to the ground trough the NASA communication infrastructure so they need to be compliant with all of NASA protocols.
- Data corruption continuously verified
- Network of Computing Centers are supporting the experiment for data reconstruction and Monte Carlo production

Summary (3)

- CNAF is supporting AMS activities with ~8000 HS06 and ~1.5 PB of disk space
- A new integrated solution for remote data access has been implemented between the ASDC and CNAF.