Data Challenges in Astronomy and Astroparticle Physics

G. Lamanna

IHPCSS, Ljubljana 28 June 2016
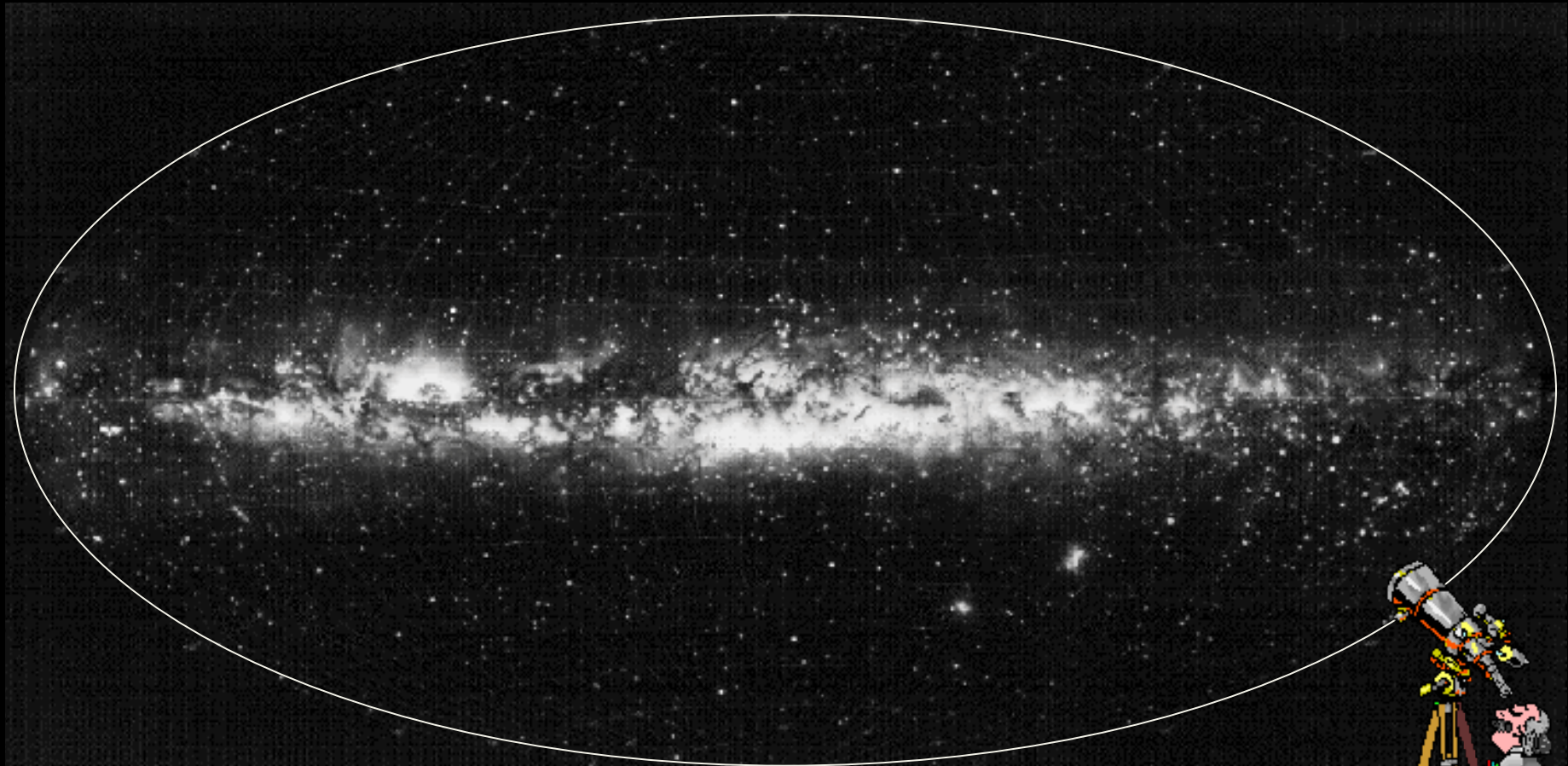
Radio   Infrared   Visible light   X rays   Gamma rays
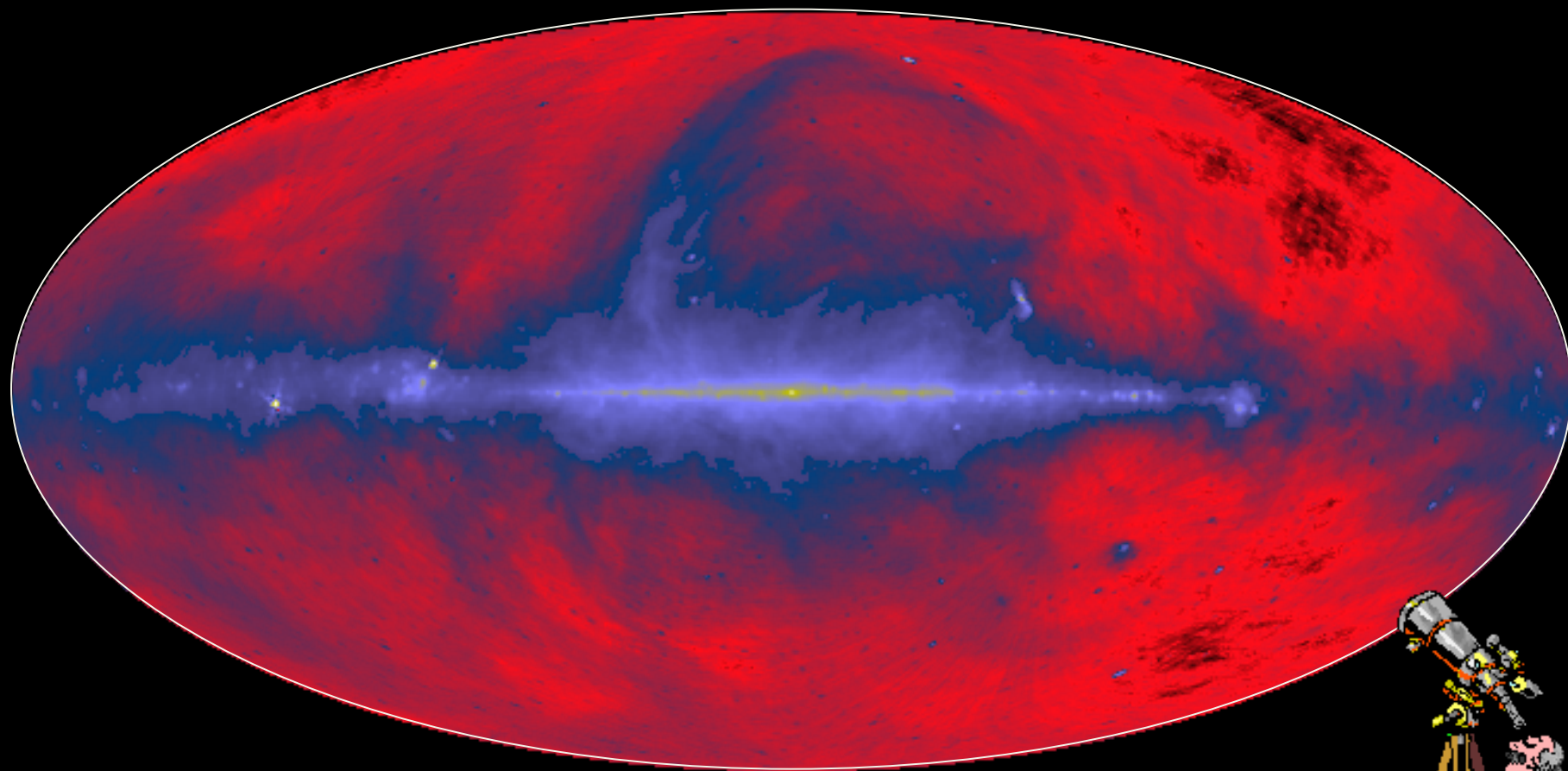   (eV)

Radio
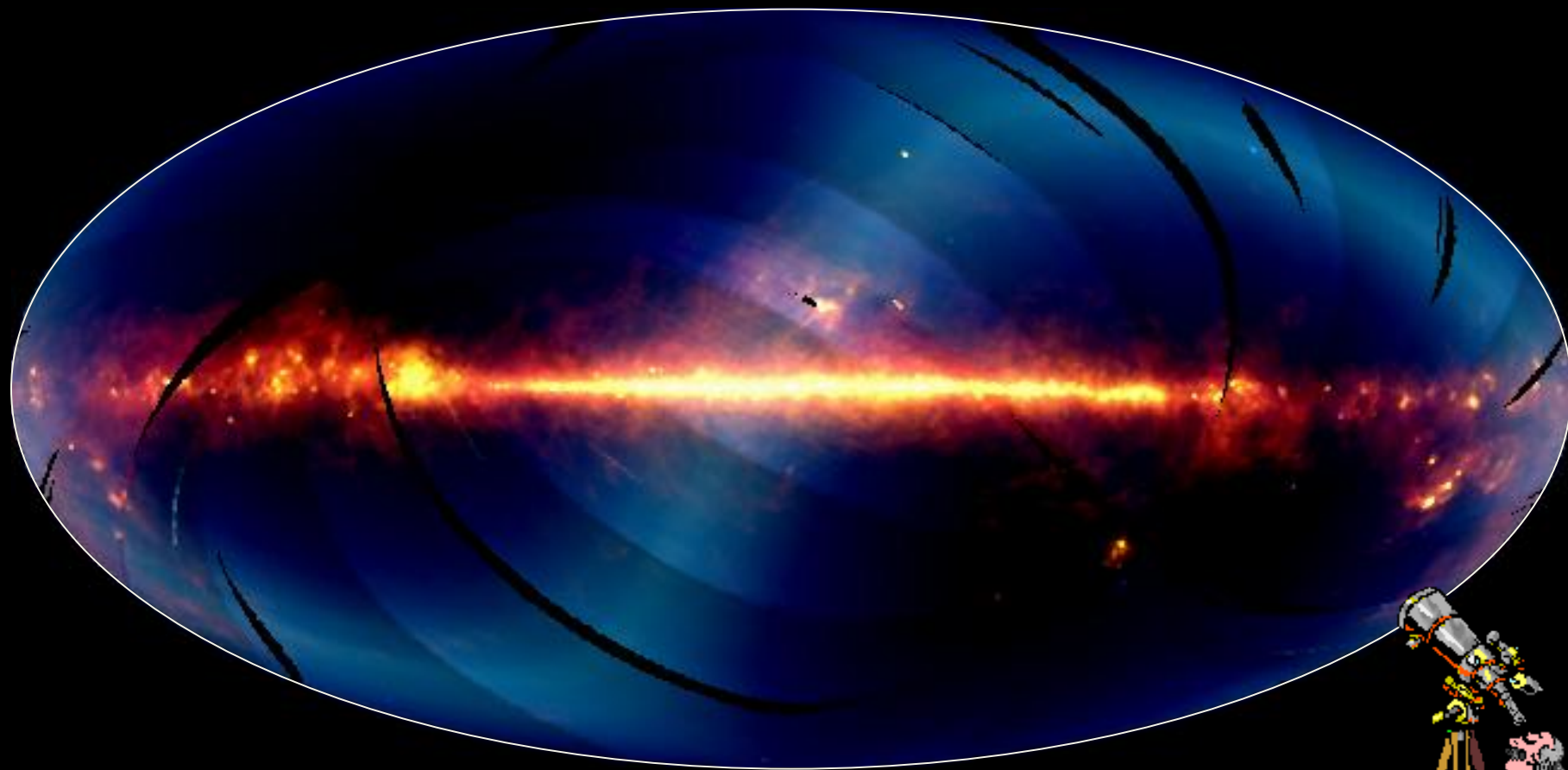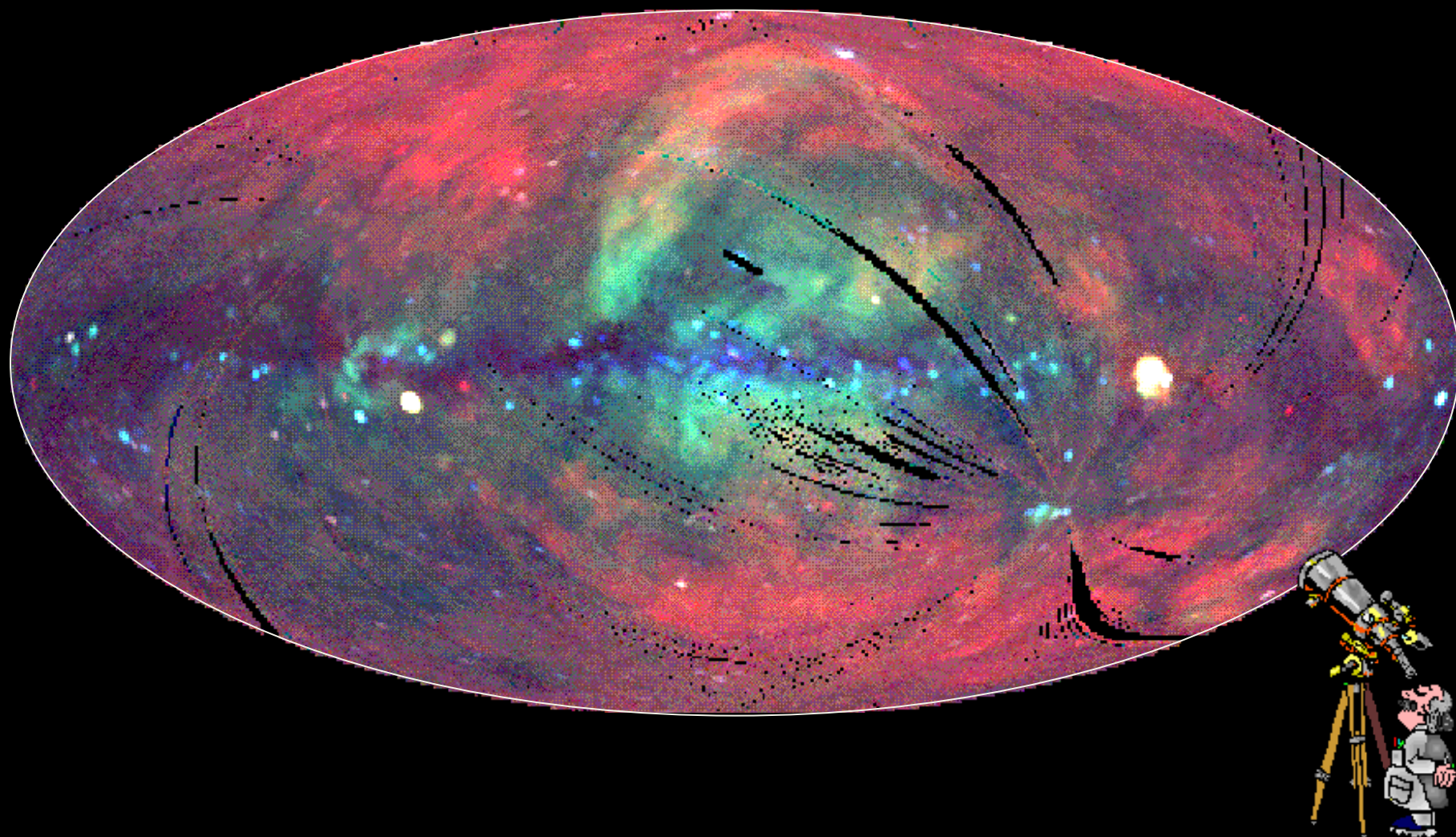$(10^{-6}$ eV)   Infrared   Visible light   X rays   Gamma rays

Radio   Infrared   Visible light   X rays   Gamma rays
          (10^{-2} eV)

Radio   Infrared   Visible light   X rays   Gamma rays
$(10^3$ eV)
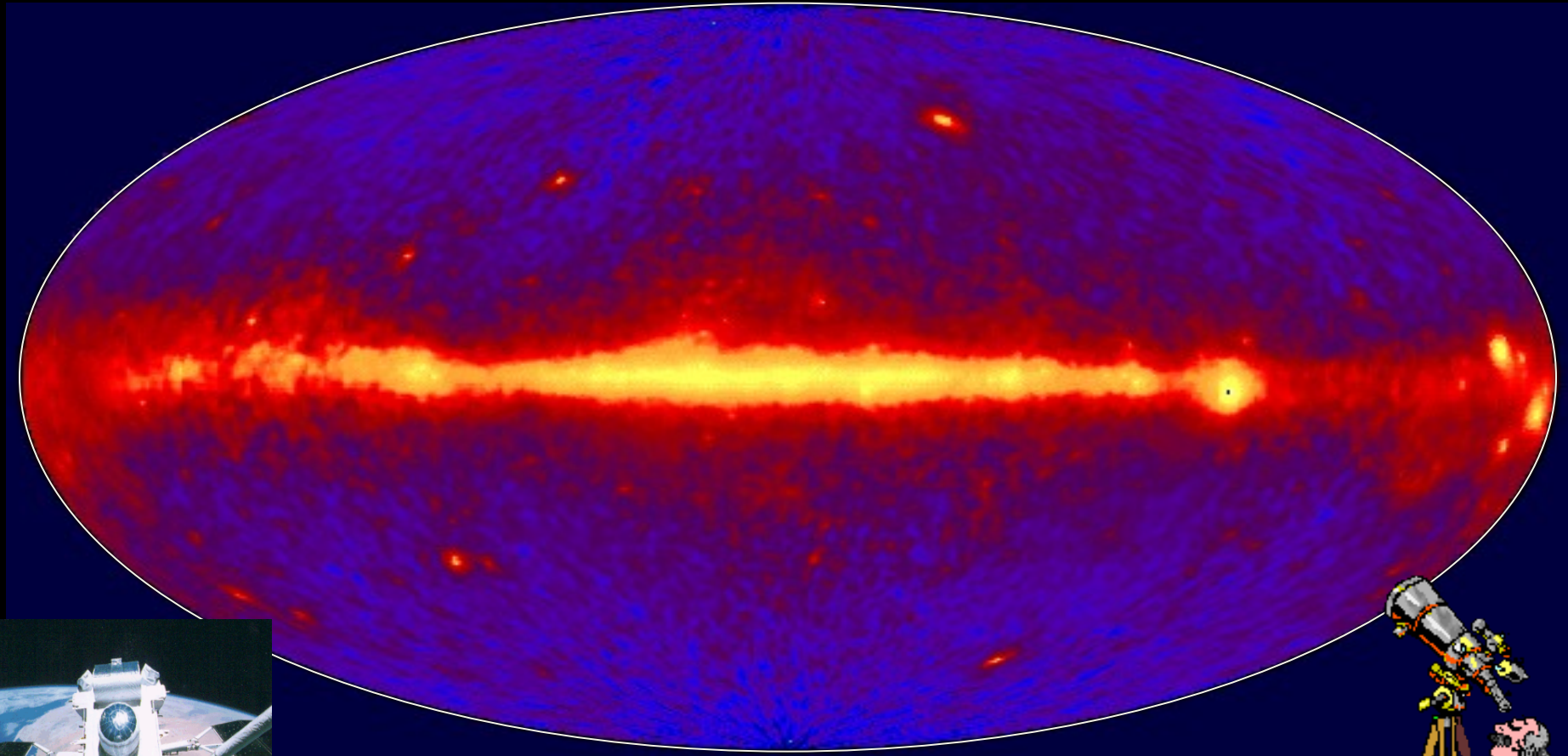
Radio   Infrared   Visible light   X rays   **Gamma rays**
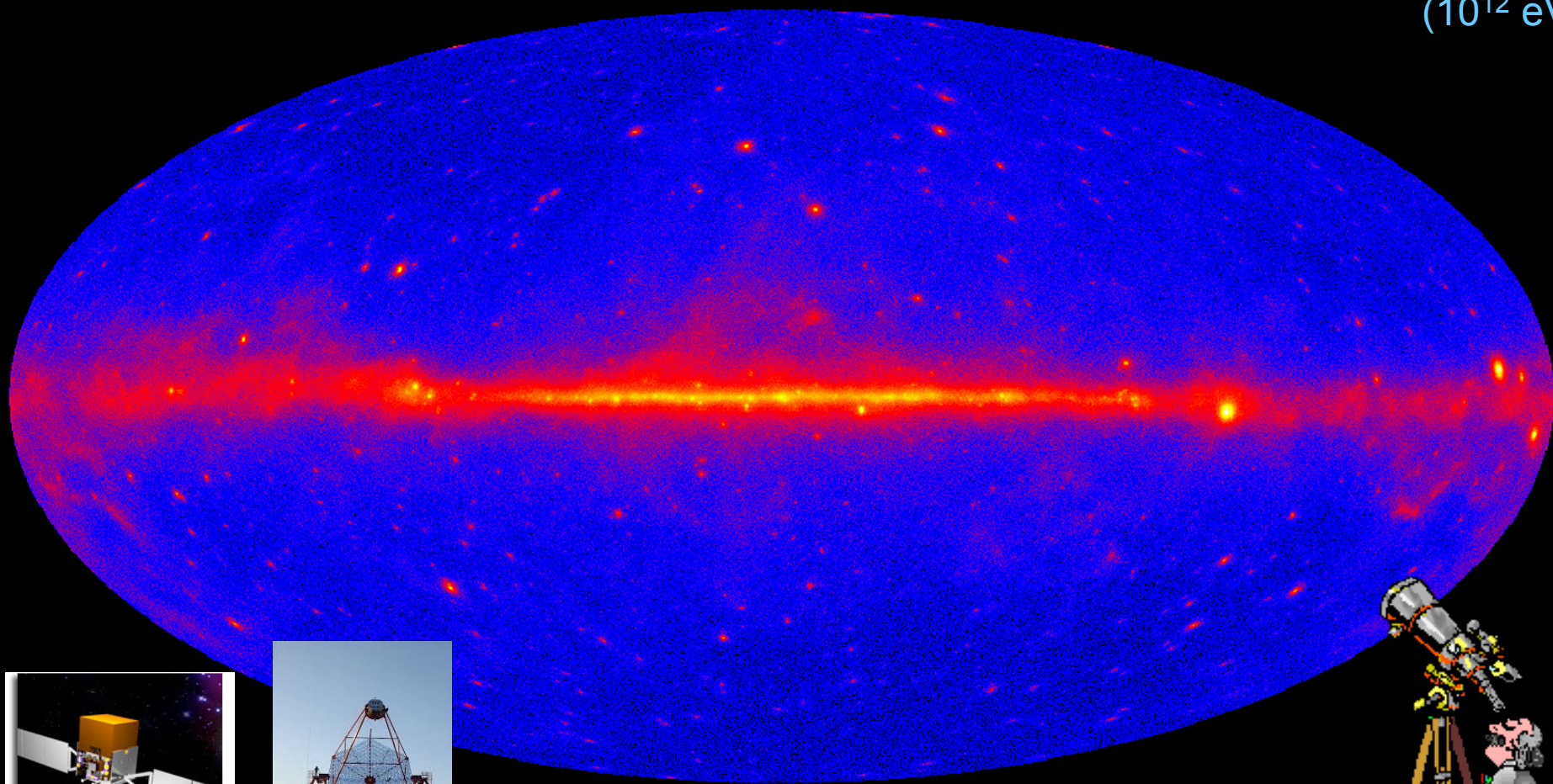**(10^9 eV)**

Radio   Infrared   Visible light   X rays   Gamma rays
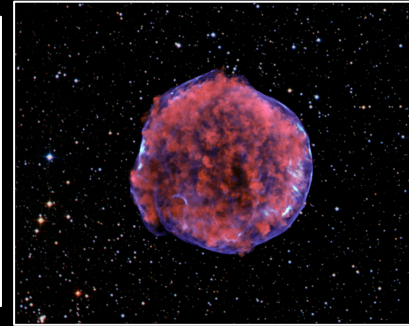$(10^9$ eV)

VHE
$(10^{12}$ eV)

Cosmic rays

Weak lensing

Pulsars
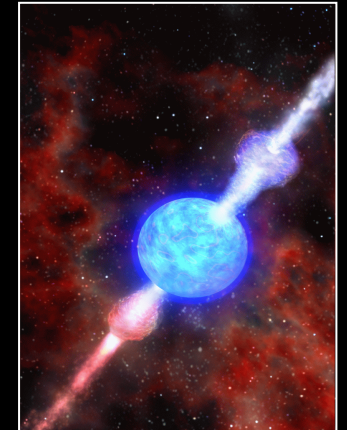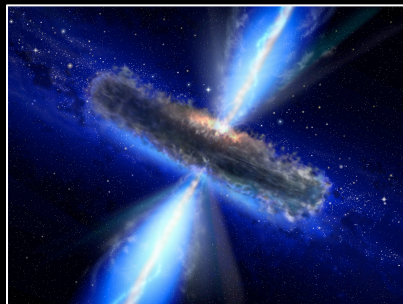
SNR

Dark energy

Dark matter

Gamma-Ray Bursts

Star Forming Regions

AGN

Binary Systems
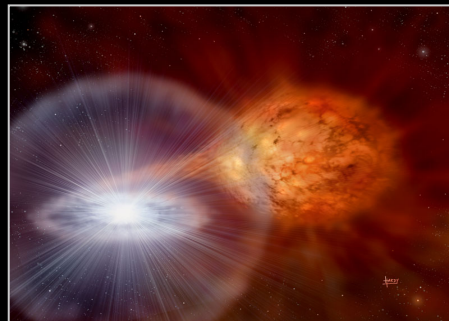
During the last decade:
- intensive construction of large astroparticle physics experiments and detectors.
- new perspectives in Astronomy and new infrastructures in preparation.

Towards the end of the decade:
- the projects passed from the noise hunting regime to the generation of large sets of data. Data production needs large computing resources, intensive simulation and large storage space.
- multi-messengers data need formats, software and services for wide accessibility and effective mining.

(Astro-)Particle Physics community tradition:

- High Throughput Computing Distributed centres for data archive, access, reduction and analysis.
- Large scientific collaborations and capacity to support and serve large communities for cooperative data analysis.
- Computing is (exclusively) sequential but budget, environmental and project specifications push towards parallelization, vectorization and HPC exploitation.

Astronomical community tradition:

- A pioneer of scientific data sharing through the Virtual Observatory: a global data infrastructure for data mining, preservation and interoperability.
- Scientific data sharing is currently a "must", but Astronomy has been always an Open data (in general after a proprietary period) research domain !
- Common multi-wavelength data formats and data tools are developed.

Astronomy and **Astroparticle Physics** domains are experiencing a **deluge of data** with the **next generation** of facilities prioritised in the European Strategy Forum on Research Infrastructures (**ESFRI**), such as SKA, CTA, KM3Net and with other world-class projects. The concerned scientific communities are committed to finding innovative solutions to solve their challenges of **big data access**, **management**, **interoperability** and **high performance computing data processing optimization**. **Large international cooperation** framework and potential **inter-disciplinary teams** are established in order to successfully face these new challenges.

Astronomy ESFRI & Research Infrastructure Cluster
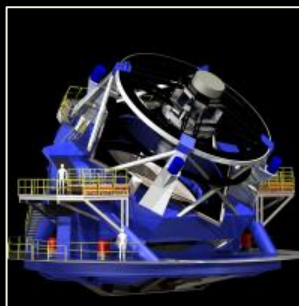*ASTERICS - 653477*

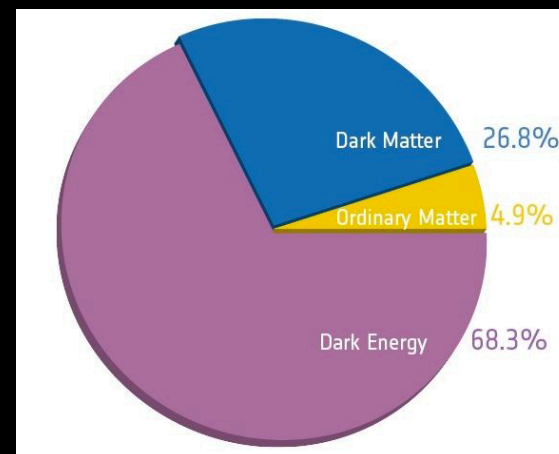**Radio**   Infrared   **Visible light**   X rays   **Gamma rays**

LSST

A Cosmology machine

- What is the mysterious dark energy that is driving the acceleration of the cosmic expansion?

- What is dark matter, how is it distributed, and how do its properties affect the formation of stars, galaxies, and larger structures?

- How did the Milky Way form, and how has its present configuration been modified by mergers with smaller bodies over cosmic time?
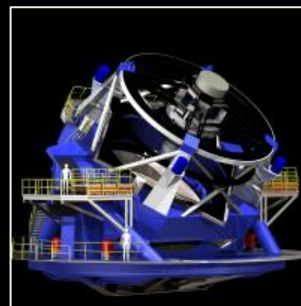
- […]

Radio　　Infrared　　**Visible light**　　X rays　　**Gamma rays**

LSST

3.2 Gigapixels camera
(six optical band)

15 TB/night data

Gigantic catalog
DataBase (50 PB)

Cumulative 500 PB
for imaging in
10 years

Deep, Wide, Fast, Optical Sky Survey
— 8.4m telescope
— Resolution: 0.2 arcsec/pixel
— 2 x 15sec exposures / 2sec readout and repositioning in 5 sec
— Location: Cerro Pachon, Chile
— Operation in 2022
— Expected computing peak ~ 1.6 petaFLOPS
— Transient Science: alerts pub. in 60 sec.

## Generated data for user access:
— ~50 galaxies detected per arcmin^2
— ~37 billion "objects"
— ~30 trillion "detections"
— Largest table: ~5 PB
— Tallest table: ~50 trillion rows

Source: J. Becla, SLAC

~ 1/1 000 000 000 of LSST data
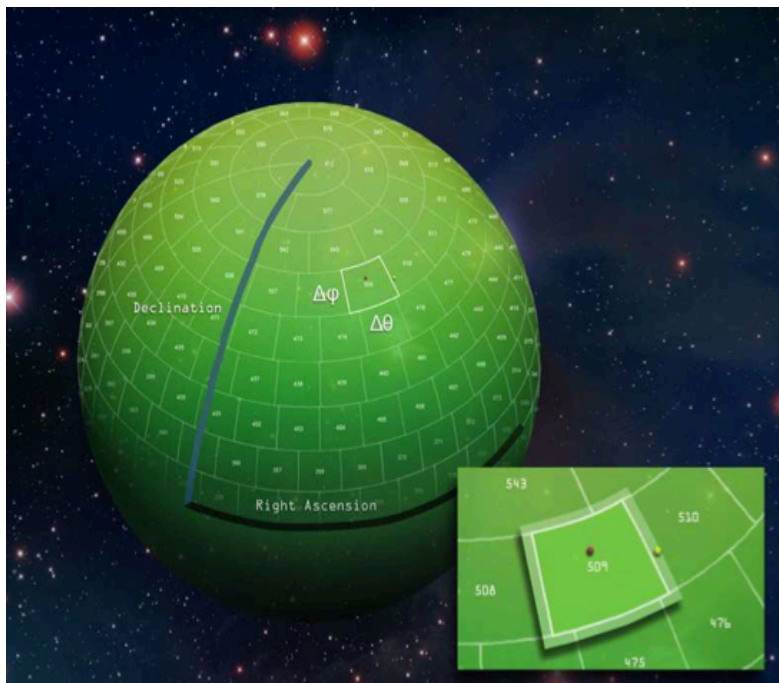

ANALYSIS REQUIREMENTS

- In a region
- Across entire sky
- Analysis of objects close to
  other objects
- Analysis that requires special grouping
- Time series analysis
- Cross match & Anti-cross match with external catalogs
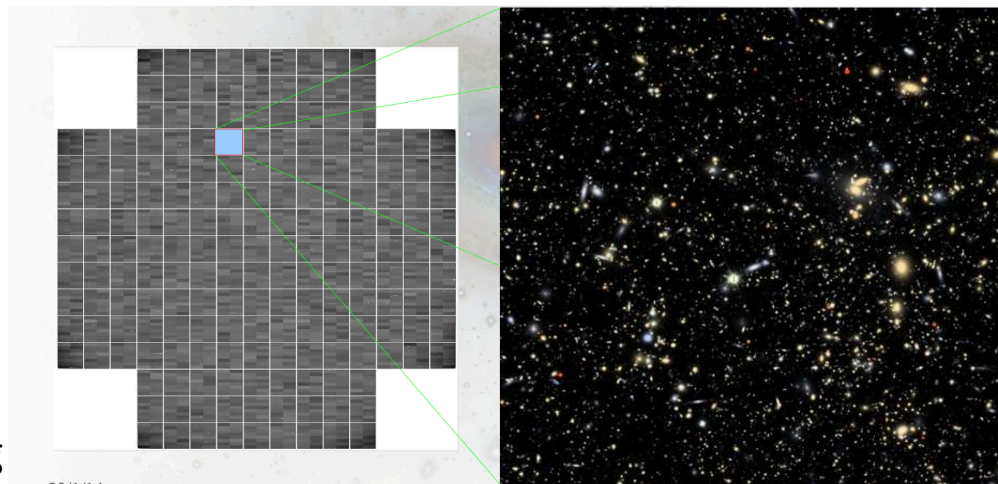
USER REQUIREMENTS

- Scans trough petabytes, multi-billion/trillion row table joins
- ~100 interactive queries + ~50 complex simultaneously, seconds-few hours
  response time !
- Complex analytics from trivial interactive image scan to spatial correlation
  and time series analysis.
- Data and software preservation over multi-decade

*Source: E. Gangler, CNRS; J. Becla, SLAC*

DESIGN
- Relational database, spatially-shared with overlaps
- Map/reduce-like processing
- Hive/HadoopDB not in the specifications
  (time dependent on data size; max size limited;
   indexing, partitioning and cache not adapted)

SLAC developed Qserv solution:
Based on existing components,
with custom glue (100% open source).

Source: D. Boutigny, CNRS
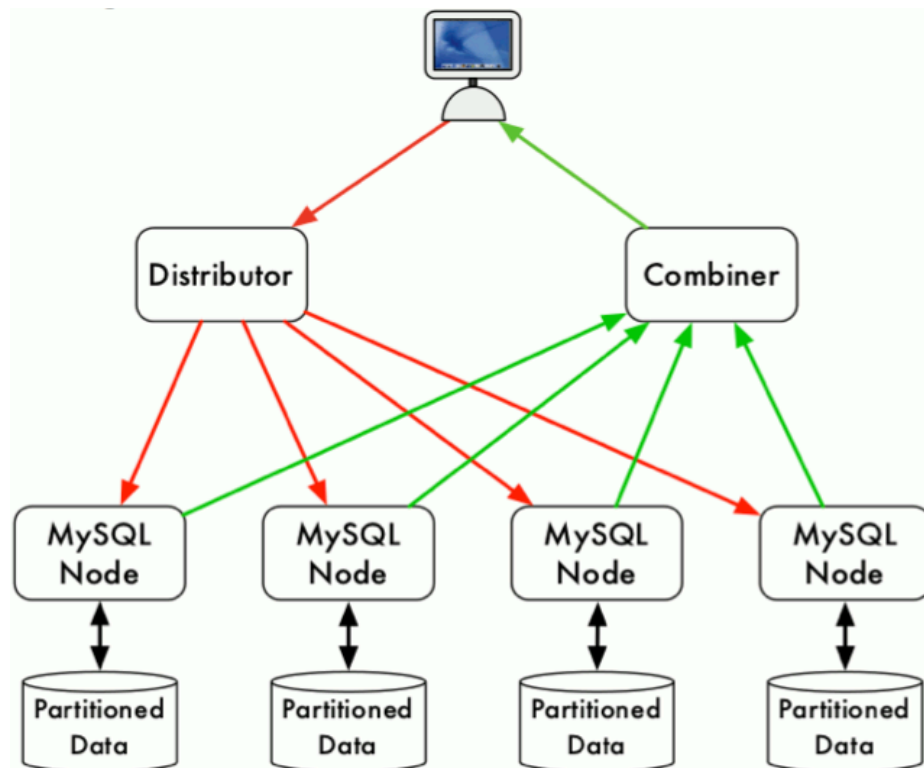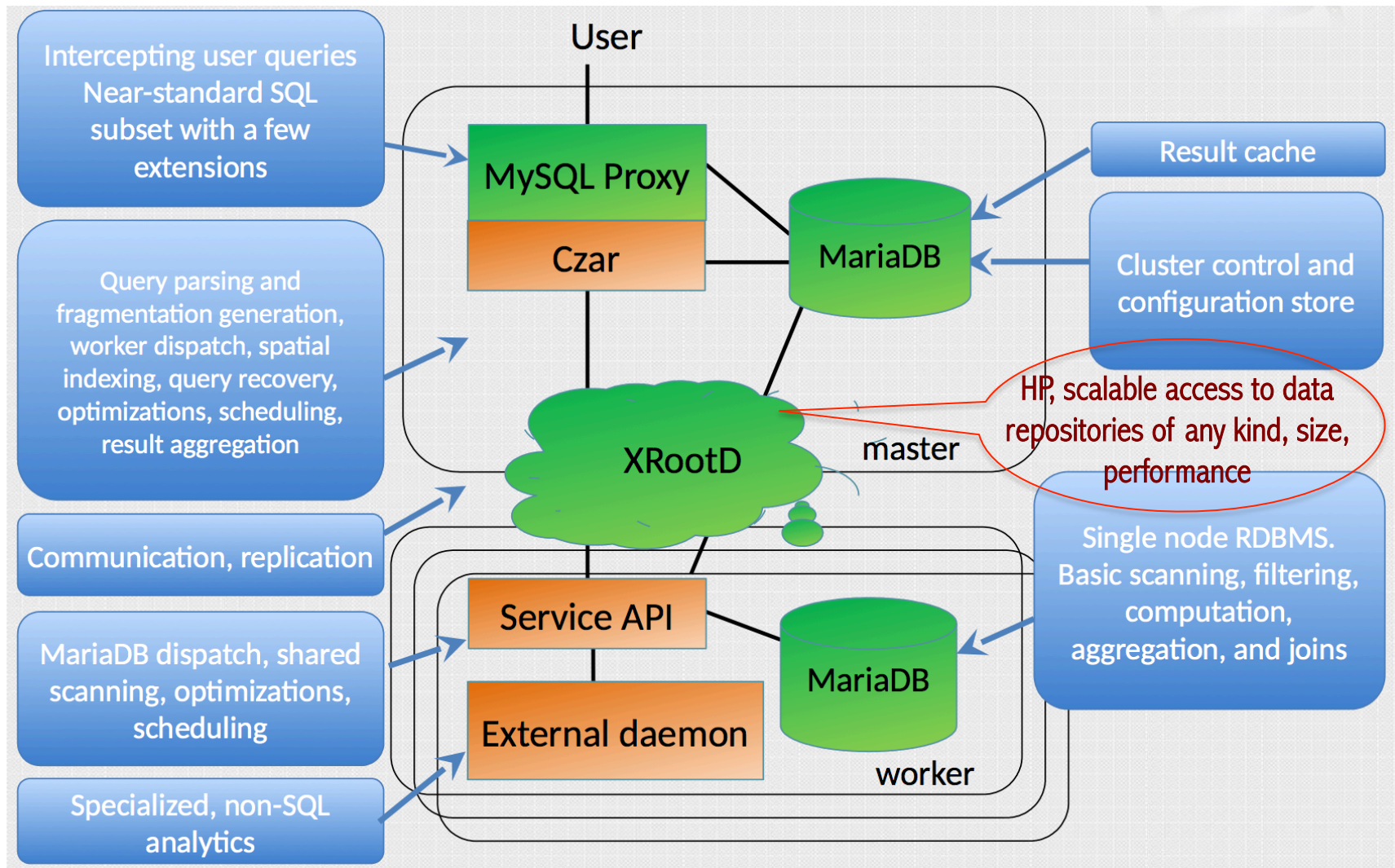
Current tests (with 50 nodes) results: 30-query scan in 5m27s; avg speed for a single query: 3m

Source: D. Boutigny, CNRS
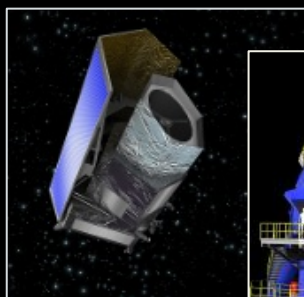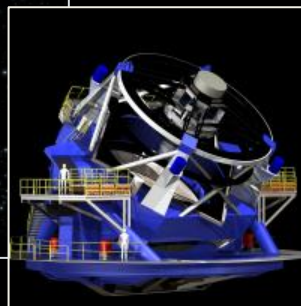
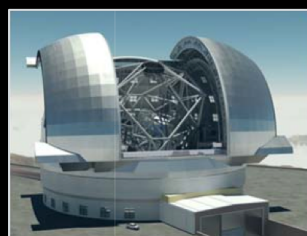**Radio** Infrared **Visible light** X rays **Gamma rays**

EUCLID

LSST

E-ELT

Radio   Infrared   **Visible light**   X rays   **Gamma rays**
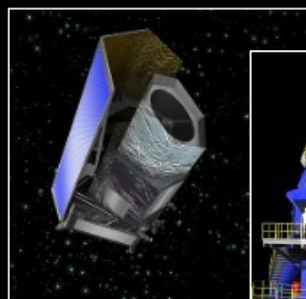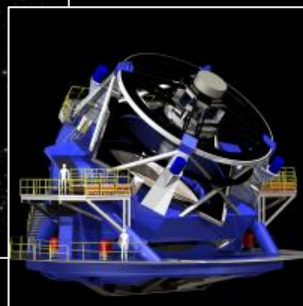
EUCLID

LSST
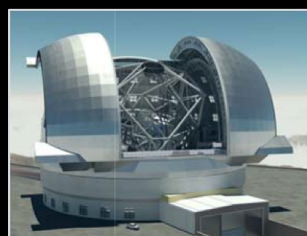
E-ELT

HESS
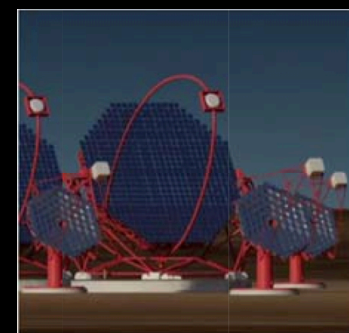
CTA

CTA
- The next generation ground-based very high energy gamma-ray instrument.

- An open observatory to a wide astrophysics community.

- It will provide a deep insight into the non-thermal high-energy universe.

CTA aims:
- ✓ Understanding the origin of cosmic rays and their role in the Universe

- ✓ Understanding the nature and variety of particle acceleration around black holes

- ✓ Searching for the ultimate nature of matter and physics beyond the Standard Model

Radio   Infrared   **Visible light**   X rays   **Gamma rays**
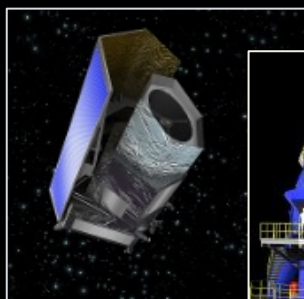
EUCLID

LSST
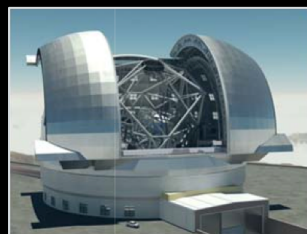
E-ELT

HESS
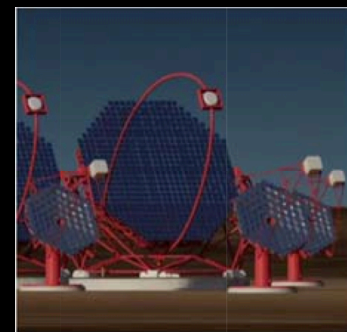
CTA

8 GB/s data rate
4 PB/year reduced raw data
Observatory: open access
to archived data

- ~ 100 Telescopes in 2 arrays (Chile and Canary).
- Three telescopes size.
- Construction starting in 2017.
- Full operation in 2022.
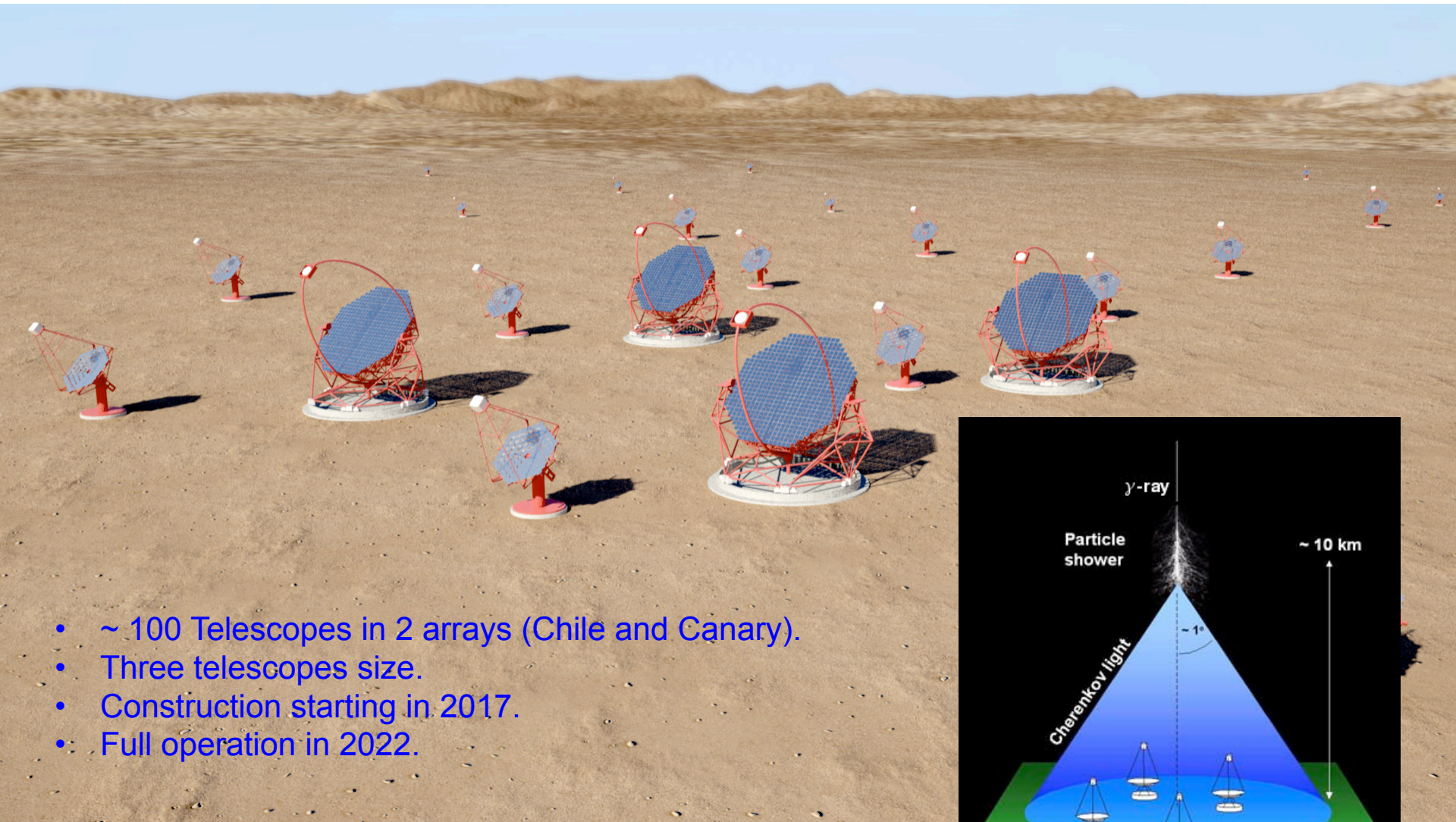
- ~ 100 Telescopes in 2 arrays (Chile and Canary).
- Three telescopes size.
- Construction starting in 2017.
- Full operation in 2022.
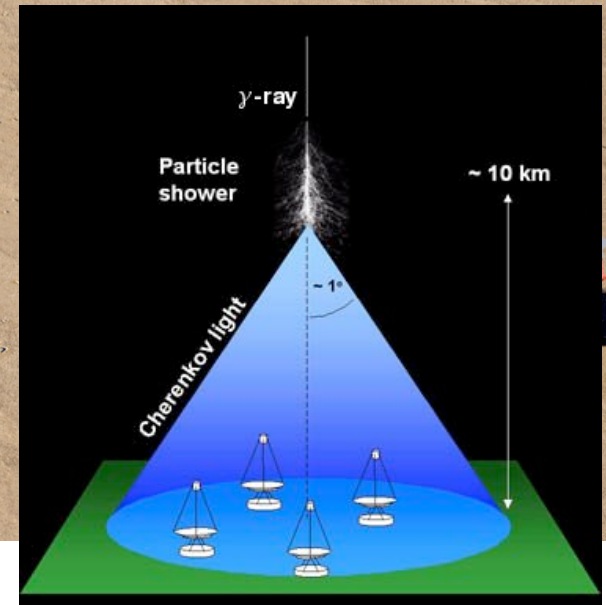
A drawer with 16 pixels (photomultipliers) detectors of a CTA camera
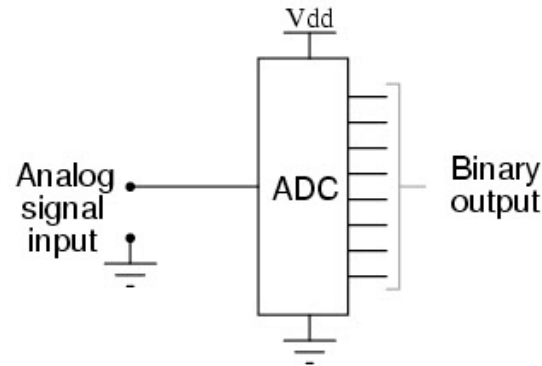
An analog-to-digital converter (ADC) to convert the voltage continuous physical quantity registered by a photomultiplier to a digital number that represents the voltage's amplitude.
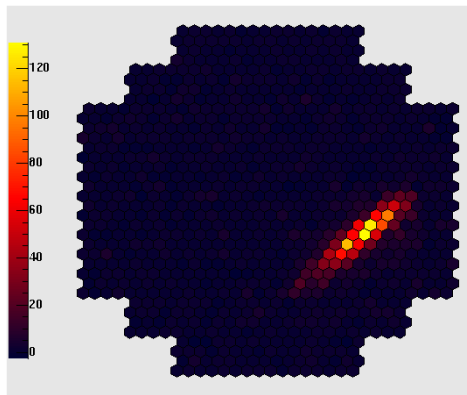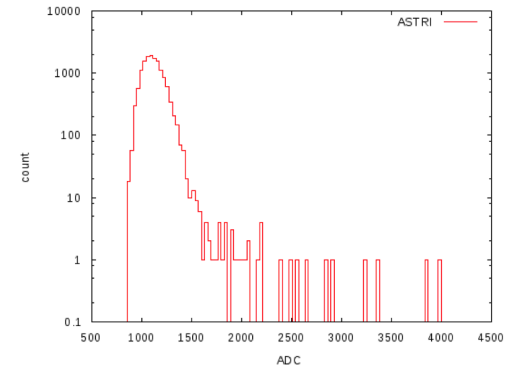
From a camera image
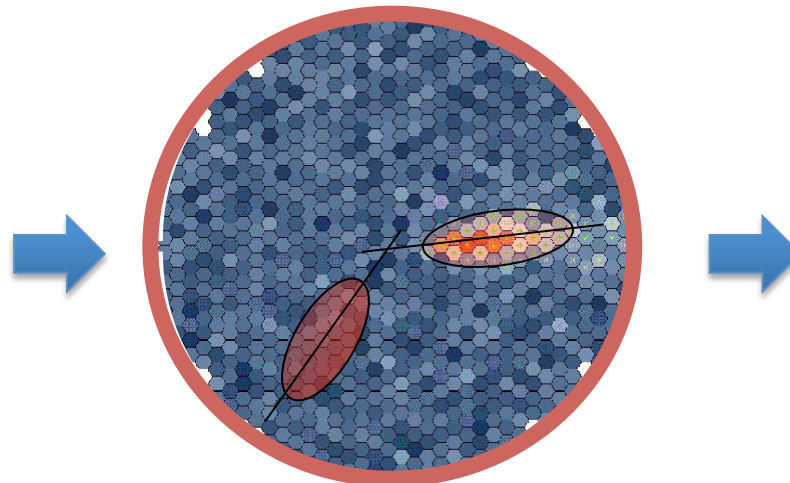
To the stereoscopic reconstruction for energy and direction determination
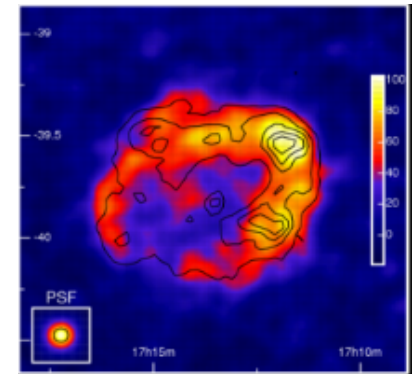
To sky images

Computing model mainly based on distributed high-throughput computing
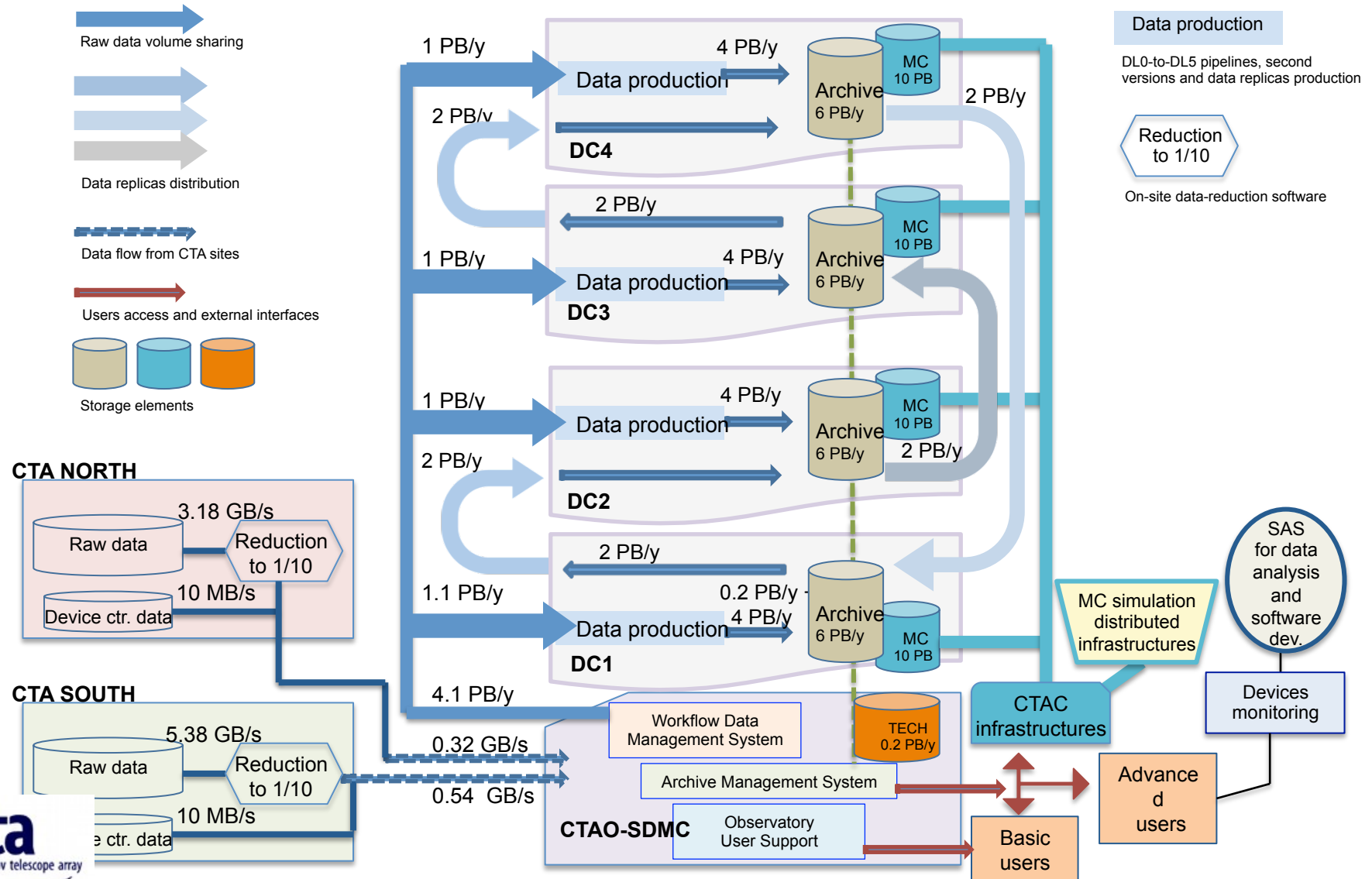
- **Need a three-pronged approach**
- **Need a combination of different domains experts**
- **Developments and tests for comparison and validation**

High Performance
Hash functions
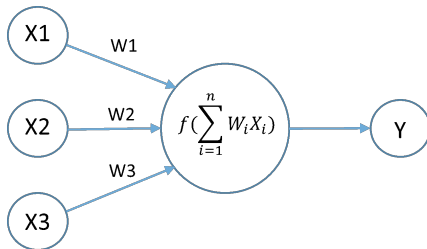
Pursue innovative approaches to increase sensitivity and cutting processing times
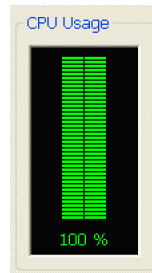
Look at accelerators and data vectorization/reduction for reducing time and costs.
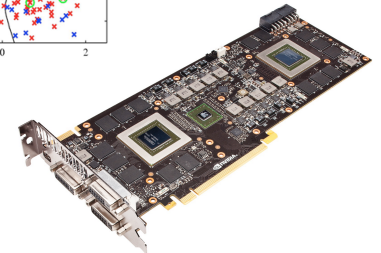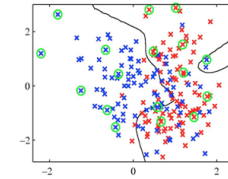
Combining algorithms and novel hardware to test computing-model modifications

A few examples of current investigation of HPC solutions within the CTA pipeline.

1) Optimization of raw data format and improved compression-speed through vectorization.

Raw data files are a series of ADC counts which together follow a typical distribution. They are stored each as unsigned int.

Vectorized Polynomial approach:

Basical idea

Store several ADC values in one single unsigned int
With a basis B (range of noise) :

$$s = b_0 + b_1 B + b_2 B^2 + \cdots$$

$$= \sum_{i=0}^{N} b_i B^i$$

Results about raw data reduction ratio and speed

	Reduction ratio	Time
LZMA (7z)	4.84	7 min 48.636 s
Basic polynom	3.65	3.671 s
Polynomial reduction + LZMA	4.65	24.341 s

Distribution

Reduced table
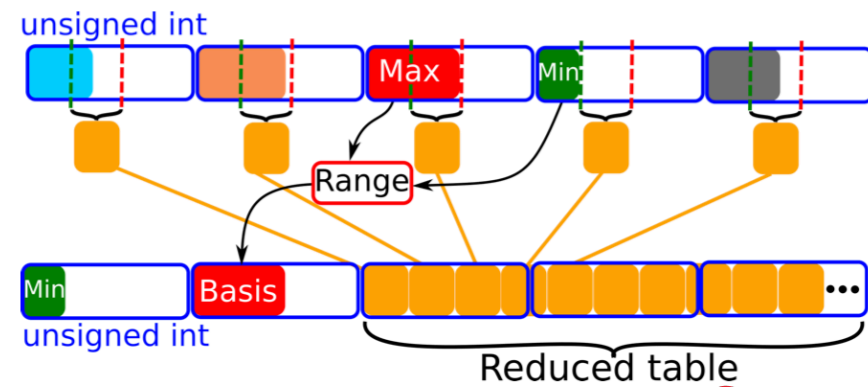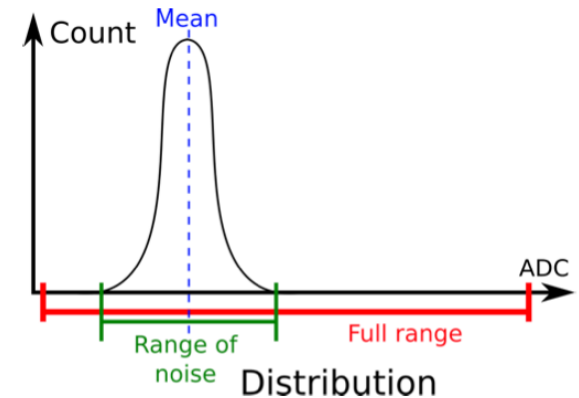
What is the good vector size?

- The size of the vectorial register
  - Yes, but which one : SSSE3, AVX2, AVX512-2 or AVX1024-2

Vectorized Polynomial Compression Without B splitting (assuming N=3).

$$N = \left\lfloor \frac{\ln\left(2^{32}-1\right)}{\ln B} \right\rfloor = 3$$

Parallelism at the level of a single CPU (the single core occupancy is optimized by this data format) and as a function of the CPU architecture.

$$s = \sum_{i=0}^{N} \mathbf{b_i} B^i$$

Be independent of Vectorial Register size when using different processors (SSSE3, AVX-2, AVX512-2, AVX1024-2).

The compressed data are not in the same order
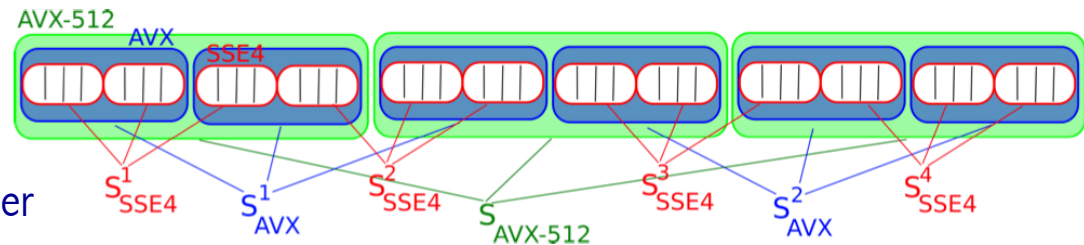
(work in progress)

With $\mathbf{v} \in \mathbb{R}^m$ (the data to compress), we have :

$$\mathbf{s}_{SSE4}^{1,0} = \mathbf{v_0} + \mathbf{v_4}B + \mathbf{v_8}B^2$$
$$\mathbf{s}_{AVX}^{1,0} = \mathbf{v_0} + \mathbf{v_8}B + \mathbf{v_{16}}B^2$$
$$\mathbf{s}_{AVX-512}^{1,0} = \mathbf{v_0} + \mathbf{v_{16}}B + \mathbf{v_{32}}B^2$$
$$\neq \quad \mathbf{s}_{AVX}^{1,0} \neq \mathbf{s}_{SSE4}^{1,0}$$

A few examples of current activities about HPC solutions implemented in the CTA pipeline.

2) Vectorization of data processing calibration + reconstruction.

Every pixel in a camera is treated independently: obviously vectorized (contiguous data, no back dependencies) . Non parallelizable because the Cherenkov image is too small in size.

- Cleaning through recognition of neighbours most lighted pixels.
(Hexagonal camera does not enable vectorization)
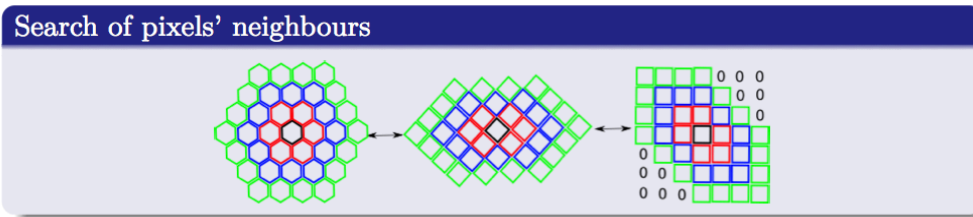
- Picture converted in diagonal matrix
(cleaning through vectorization is enabled)

Search of pixels' neighbours

Calibrated picture
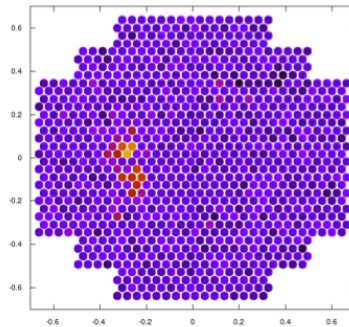
Picture converted in matrix
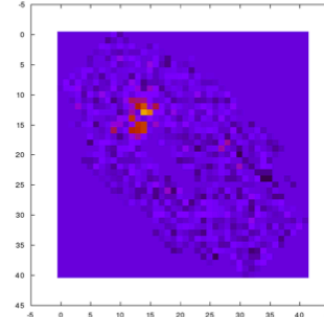
Cleaned picture
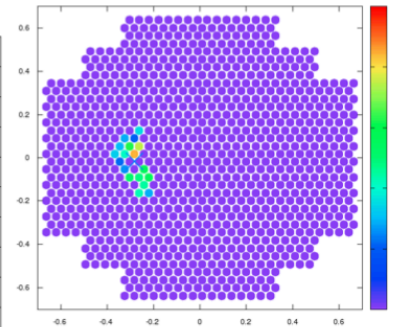
Optimized HPC data format.

Enabling CPU data prefetching
- With contiguous data
- Guaranty data locality
- Cache friendly

Enabling vectorization
- With aligned data for vectorizable computation
- Of algorithms: calibration, cleaning, Hillas gamma-ray event reconstruction

Full consistent description:
- All telescopes array
- All cameras
- All possible layout
- All reduced level of data

Before pixels sorting
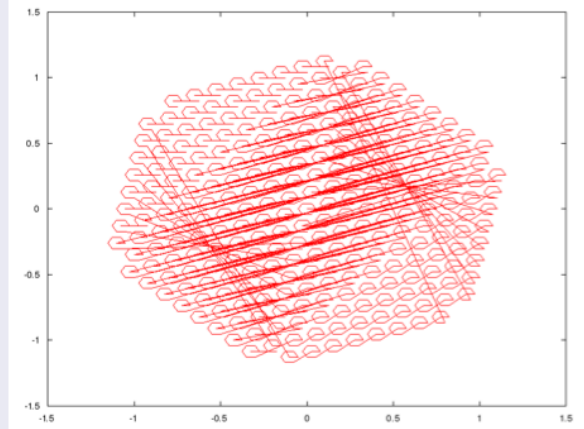
After pixels sorting
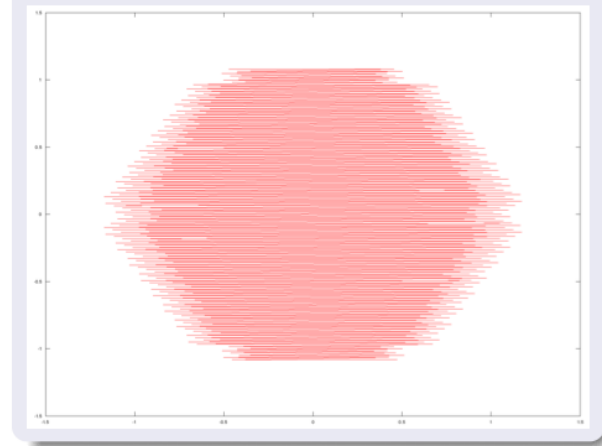
It allows faster pixel's neighbour search and cleaning .

Hillas parameters calculation.

- Reduction : Sum of all pixels photoelectron signal per camera
- First momentum : ellipse's position
- Second momentum : ellipse's orientation
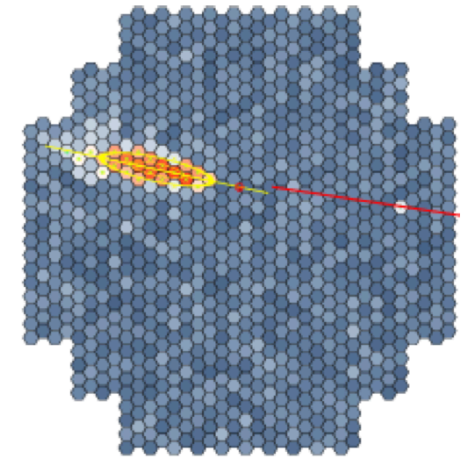
Vectorization enables important acceleration (Processing time between 9200 and 46000 evt/s per tel.)

And not limited by I/O:

Sum of all pixels photoelectron signal per camera

	Speed (cy/el)	Speed up
Classical	2.69842	1
Vectorized (GCC, SSE4)	0.702845	3.8
Intrinsics Vectorized SSE4	0.226675	11.9
Intrinsics Vectorized AVX	0.11379	23.7

Hillas

	Speed (cy/el)	Speed up
Standard data format	2125.5	1
New data format	53.1375	40.0
+ intrinsics vectorization SSE4	6.39931	332
+ intrinsics vectorization AVX	2.98499	712

Grid worker node
- GPFS file system (Disks on Storage Area Network)
- Ethernet 1 Gb/s (125 MB/s)

	GPFS	RAM
dd (read only)	115 MB/s	5.2 GB/s
Full analysis (IO + CPU)	74 MB/s	212 MB/s

A few examples of current activities about HPC solutions implemented in the CTA pipeline.

3) Parallel processing, low power hardware accelerator

Testing the NVIDIA Jetson TK1 as the platform for an embedded implementation:

- Parallelization of the algorithms.
- Build phase not depending on CUDA toolkit
- Software modules detect GPUs in the system and act accordingly (*thanks C++ polymorphism!*)
- CPU/GPU execution switchable on user request
- ARM cores for low-power operation.

Result of full data processing in 12.5s: 4400 evt/s, 1 telescope.

- ✓ 1.4x slower than separate modules
- ✓ 30x less power
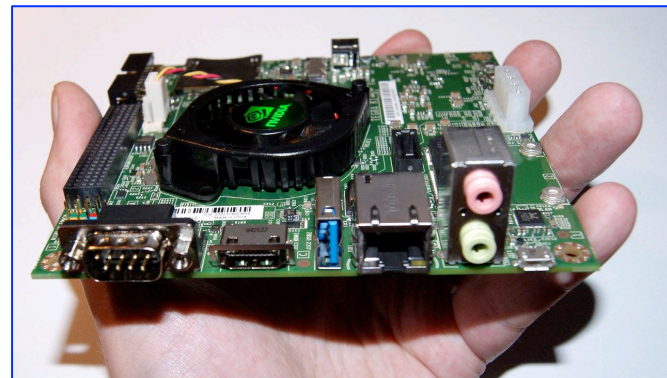
Source: A. Antonelli, INAF; D. Bastieri, U.Padova

Radio Infrared **Visible light** X rays **Gamma rays**

LOFAR

EUCLID

LSST

HESS

SKA

E-ELT

CTA
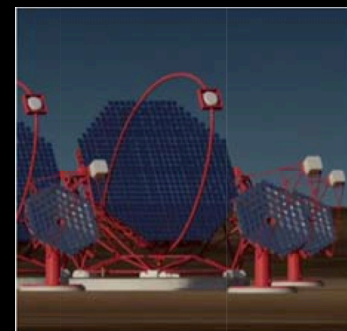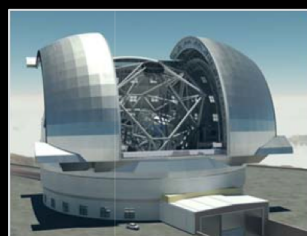
Testing Gravity
For Astrophysics and Cosmology
Radio Transients

Radio   Infrared   **Visible light**   X rays   **Gamma rays**

LOFAR

EUCLID

LSST

E-ELT

HESS

CTA
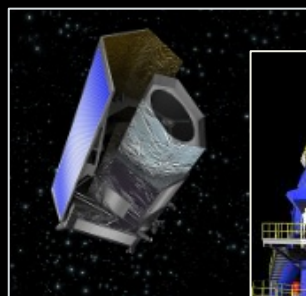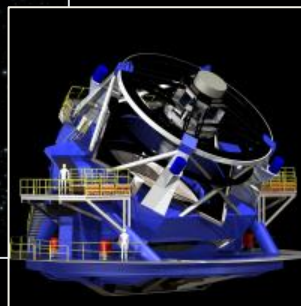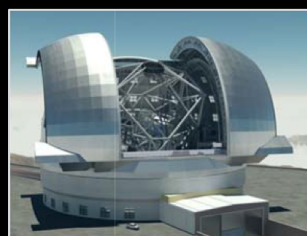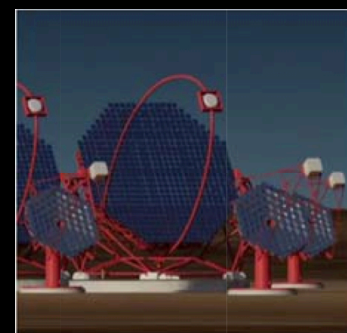
SKA

From 500 to 2000 antennas*

2 TB/s of data volume

62 EB/year

The dishes of the SKA will produce 10 times the global internet traffic

Two sites:
- Australia (*Low-Frequency Aperture Array*)
- South Africa (*Mid-Frequency Aperture Array; Dishes*)

Standard interferometer

- **Visibility:**

$$V(B) = E_1 E_2^*$$
$$= I(s) \exp(i \, \omega \, B.s/c)$$

- **Resolution determined by maximum baseline**

$$\theta_{max} \sim \lambda / B_{max}$$

- **Field of View (FoV) determined by the size of each dish**

$$\theta_{dish} \sim \lambda / D$$

Labels in diagram: Astronomical signal (EM wave), s, $B . s$, Detect & amplify, Digitise & delay, Correlate, Integrate, Process, Calibrate, grid, FFT, **SKY Image**, B, 1, 2

Source: P. Alexander, UCAM

The SKA system overview

8.8 Tbits/s

5 Tbits/s

7.2 Tbits/s

~50 PFlop

~100 PFlop

Source: C. Broekema, ASTRON
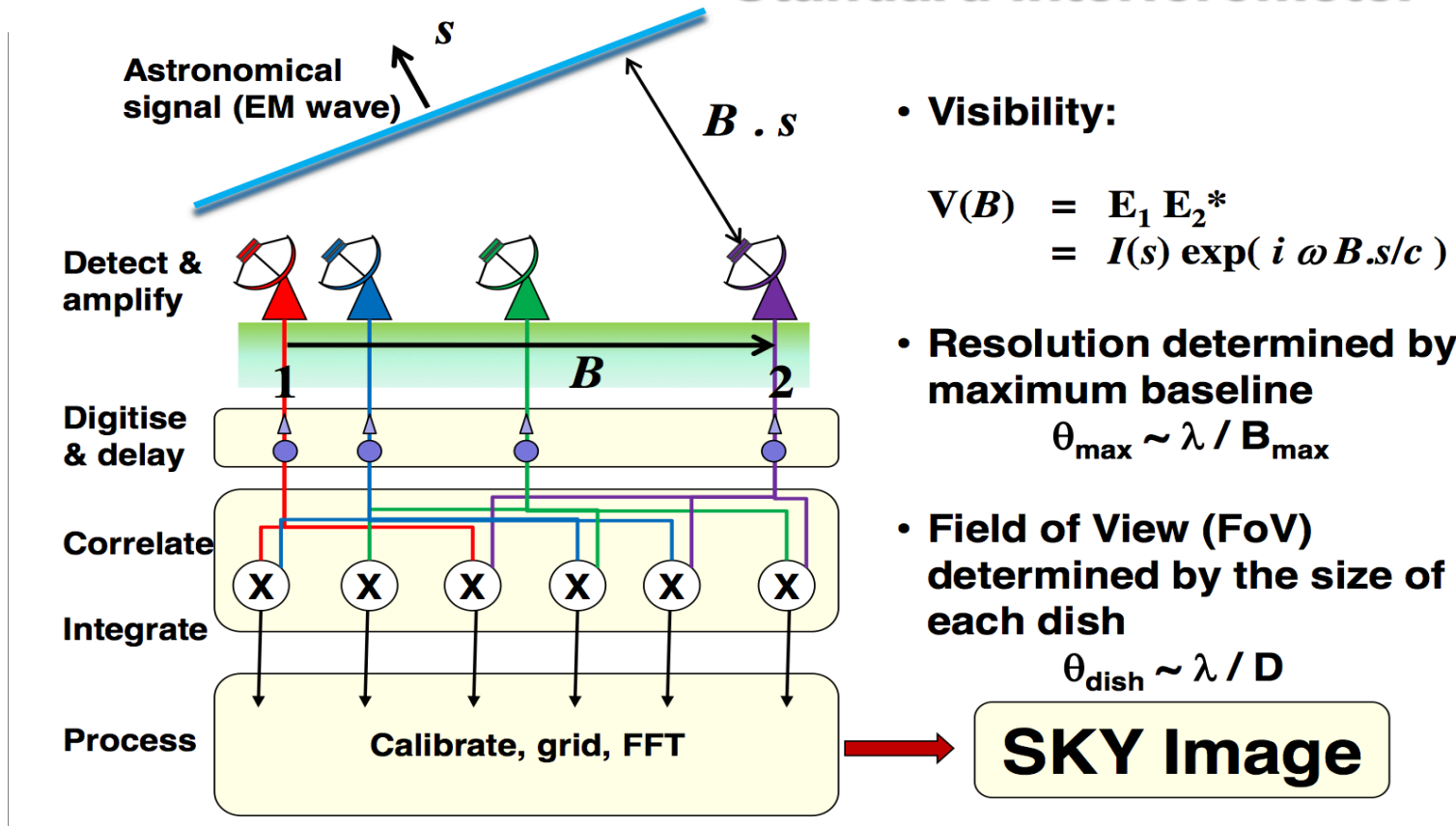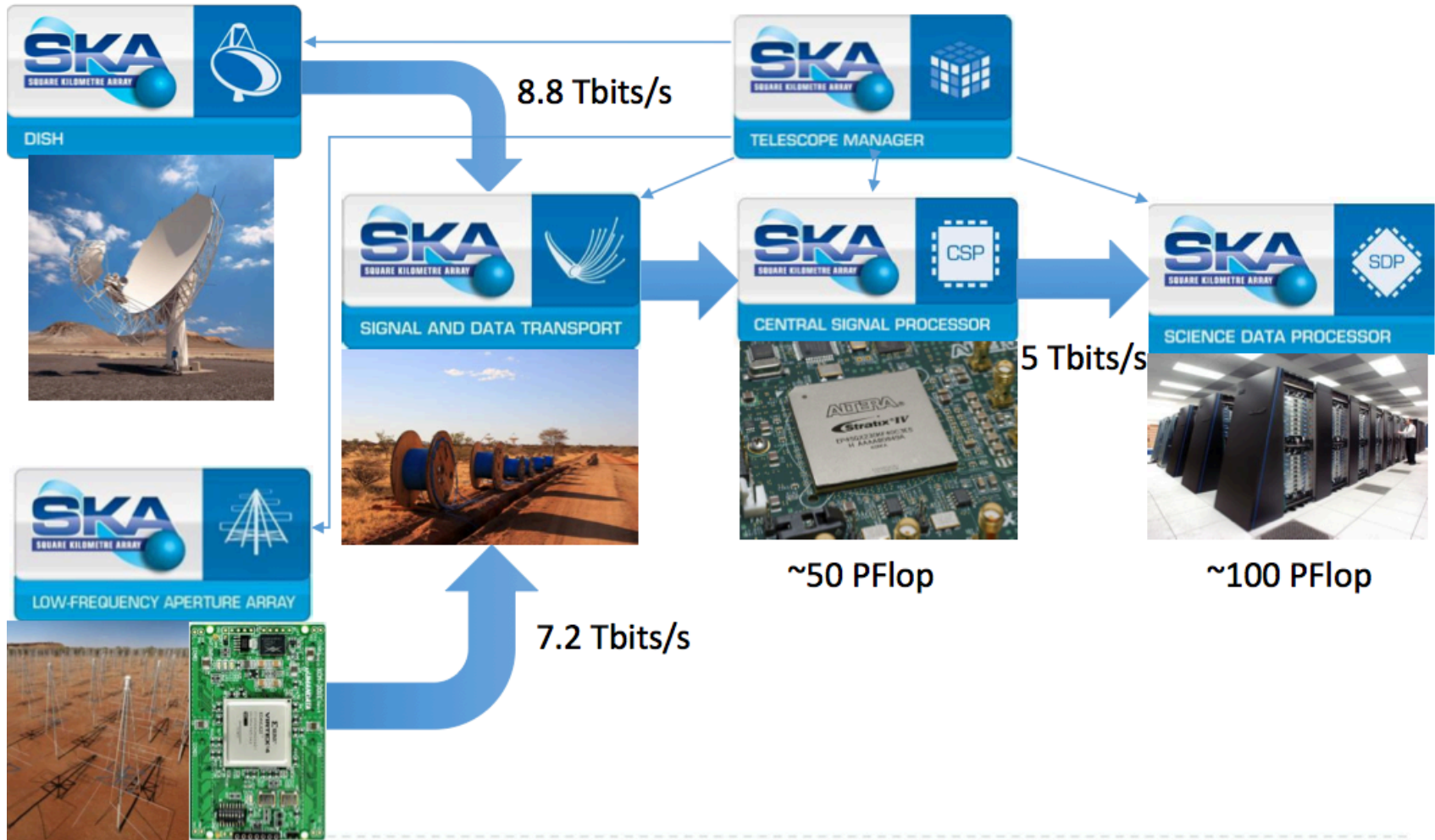
SKA is a huge computational challenge

- SKA is seen as a key programme in global IT development
- Power is also a major driver.
- CSP ~ 50 Pflop, 5 MW
- SDP ~ 100 Pflop, 5 MW
- (Tianhe-2 ~30 Pflop, 40 MW)
- Traditional HPC is not a good match because the problem is bandwidth dominated.
- Showcases a major development area of High Performance Data Analysis (HPDA).
- Software complexity is also beyond what has been achieved in astronomy previously.

Source: C. Broekema, ASTRON

eLAPP

SKA Phase 1

Telescope Manager

Science Data Processor

SDP Local Monitor & Control

CSP

1 Tera Byte/s

Data Processor

High Performance
- ~100 PetaFLOPS

Data Intensive
- ~100 PetaBytes/observation (job)

Partially real-time
- ~10s response time

Partially iterative
- ~10 iterations/job (~3 hour)

Data Preservation

High Volume & High Growth Rate
- ~100 PetaByte/**year**

Infrequent Access
- ~few times/year max

Delivery System

Data Distribution
- ~100 PetaByte/year from Cape Town & Perth to rest of World

Data Discovery
- Visualisation of 100k by 100k by 100k voxel cubes

Regional Centres & Astronomers

Source: B. Nikolic, UCAM

A major processing challenge:

To deliver SKA ICT infrastructure need a strong multi- disciplinary team

- Radio astronomy expertise
- HPC expertise (scalable software implementations; management)
- HPC hardware (heterogeneous processors; interconnects; storage)
- Delivery of data to users (cloud; UI …)

-> Approach: Co-design of software and physical layer architectures

Source: P. Alexander, UCAM; C. Broekema, ASTRON

Radio   Infrared   **Visible light**   X rays   **Gamma rays**

LOFAR

SKA

EUCLID

LSST

E-ELT

HESS
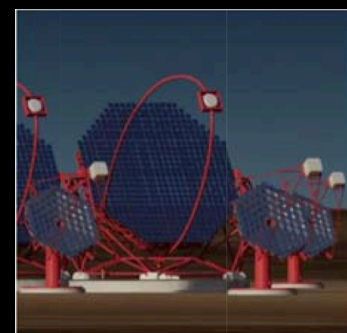
CTA

Gravitational Waves   **Neutrinos**

KM3NeT

LIGO & VIRGO   ICECUBE
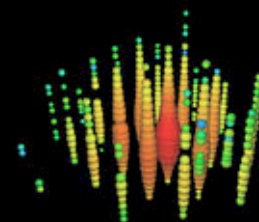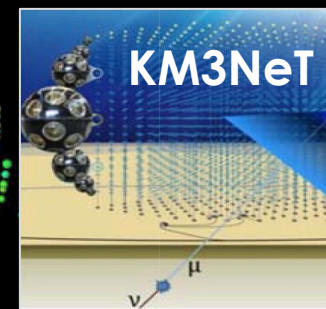
Distributed, high-throughput computing

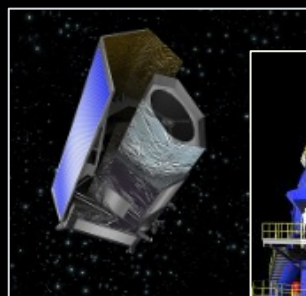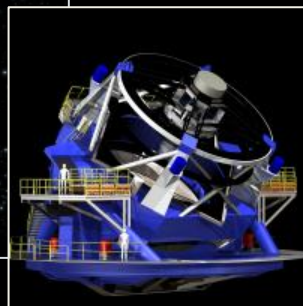- LIGO's computing infrastructure is distributed around the world.
- Low-latency computing for incoming data is distributed.
- Other workflows are less time-sensitive, and many are embarrassingly parallel.
- In addition LIGO codes run on national multi-petaflops systems funded through NSF's XSEDE program, e.g, Stampede supercomputer at TACC and the Comet system at the SDSC.
- LIGO also captures unused cycles from campus grids using NSF and OSG, as well as from volunteer computing from the Einstein at Home project.
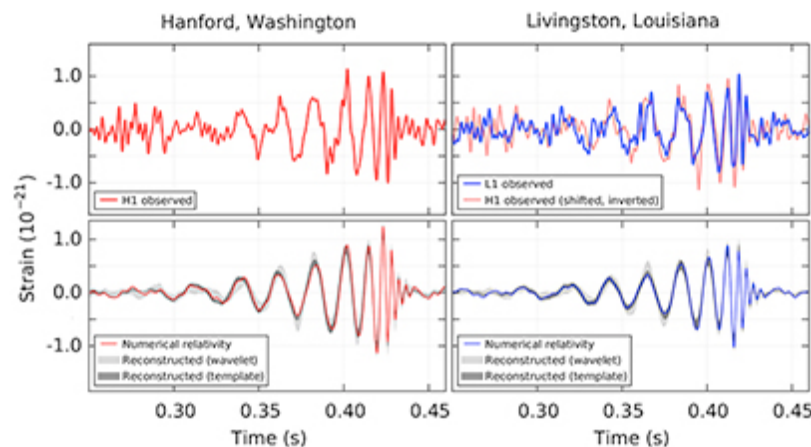
Rising requirements

- Much of the work involves single-precision, floating-point operations, with heavy use of numerical libraries including the Intel Math Kernel Library (Intel MKL) and the FFTW open source C subroutine library for computing discrete Fourier transforms

- From few PB rising to larger data sets and large number of floating-point operations needed with Advanced detectors.

ASTERICS .. is about commons and cooperation

(https://www.asterics2020.eu/)

- Astronomy ESFRI & Research Infrastructure Cluster

- Horizon 2020 Work Programme INFRADEV-4-2014/2015 Call – "Implementation and operation of cross-cutting services and solutions for clusters of ESFRI and other relevant research infrastructure initiatives"

- Focus of ASTERICS: SKA, CTA, KM3NeT, close links to E-ELT, EGO, EUCLID, LSST.

- Funded at 15 M€ for 4 years (started on 1/5/2015)

- 22 partners in 6 countries, representing a major collaboration in Astronomy/ Astrophysics/Astroparticle Physics
  ASTRON, CNRS, INAF, UCAM, JIVE, INTA, UEDIN, UHEI, OU, FAU, VU, CEA, UVA, UGR, FOM, IEEC, IFAE, UCM, INFN, STFC, DESY, SURFnet.

ASTERICS .. is about commons and cooperation

(https://www.asterics2020.eu/)

- Targeting common ESFRI-projects « Data Challenges ».

- Scopes:
  - o Enable interoperability and software re-use.
  - o Enable open standards and software libraries for multi-messenger data.
  - o Develop common solutions, share prototypes, exchange experience.

- Expected impact:
  - o Economies of scale and saving resources.
  - o <u>Contribute to the construction and operation</u> of ESFRI projects.

Working on commons along the "data flow":

1. Data generation

2. Data systems integration

3. Data analysis
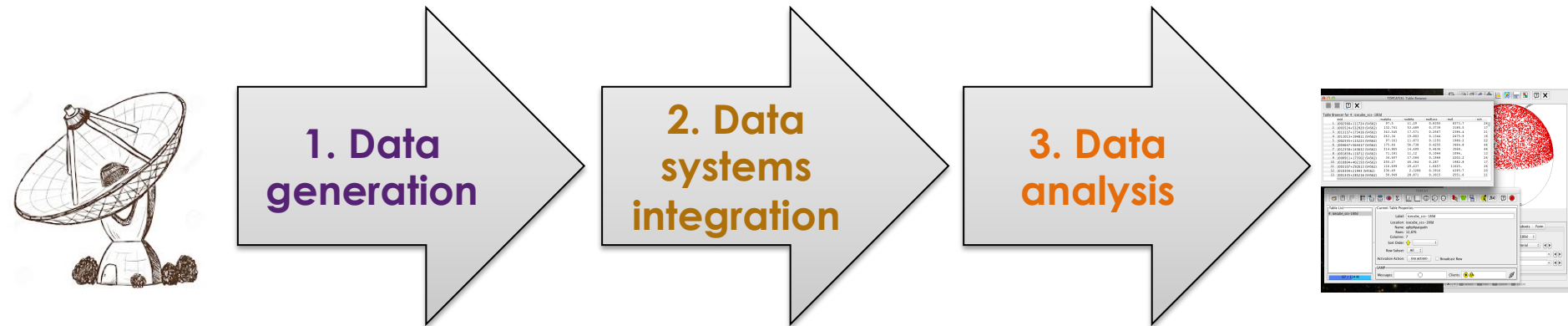
Large international partnership cooperating around three main steps of data pipelines of major ESFRI projects in Astronomy.
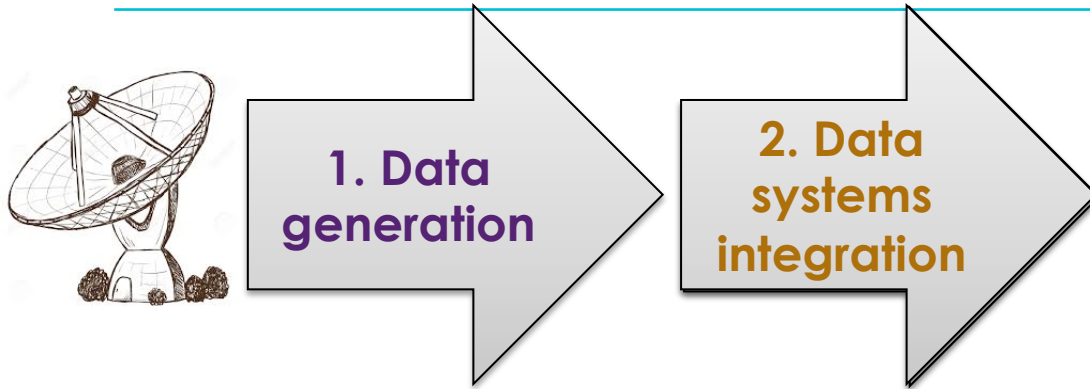
OBELICS TASKS:

1. **D-GEX: Data GEneration and information eXtraction**
2. **D-INT: Data systems INTegration**
3. **D-ANA: Data ANAlysis/interpretation**

1. Data generation

✓ Surveying real-time or close-to-detector data streaming frameworks.

(e.g. Hadoop, ACS and others; aiming at file and metadata management, fast algorithms integration, automatic remote acquisition, identification and ingestion..)

✓ Standards on data model and data format.

(e.g. Protocol buffer saving bandwidth; HDF5 simplifying big-data structure; evolution of scientific data FITS format; streaming protocols adopted in space projects...)

✓ Prototype libraries handling secondary data streams.

(environmental and engineering data, temporary local archive, device control software and observation scheduling)

✓ Benchmarking low-power computer platforms. *(GPU + ARM, FPGA, Microservers, …)*

Astronomy ESFRI & Research Infrastructure Cluster
ASTERICS - 653477

1. Data generation

2. Data systems integration

✓ Scaling-up existing databases and storage architectures beyond the Peta-scale level for complex queries.

(Cooperative activities on "identification and archiving interesting data products")

i) Developing prototype benchmarks of large size DB: Cassandra, MongoDB, Qserv.

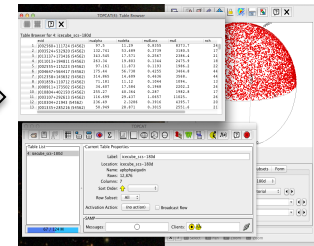*ii) Testing and adopting data-management- system services for data-sets integration:*
FLUME, RUCIO, …

iii) Multi-parameter instrument response function integration.

1. Data generation → **2. Data systems integration** → **3. Data analysis**

✓ Open source software for data analysis.

    a) *Bayesian and likelihood analyses approaches for cross-matching between catalogues and transients detected via different instruments.*

    b) *Simultaneous feature classification and extraction in multi-dimensional/multi-resolution data where the data are from multiple instruments.*

    c) *Effective likelihood reconstruction methods and new graphical processing approaches.*

✓ Workflow architectures for Peta-scale datasets on distributed computing infrastructures.

    a) *orchestration of compute intensive analysis of petascale datasets on distributed computing infrastructures (workflow engines on distributed systems, AAA protocols.)*

The (G8) political statements:
- Scientific research data should be open
- Data should be easily discoverable, accessible, useable and interoperable
- Recognition of researchers fulfilling open science principles
- Appropriate digital infrastructures.

Towards:
- Data production and dissemination challenge our research
- Data format, data services and data infrastructures should be as much common as possible.
- Training "Computing and Data Analysts".
- A global Open Science Cloud of data repository.

Astronony and Astroparticle Physics new world-wide projects
are challenging these perspectives:

• They are participating to the multi-disciplinary and multi-messenger research
  environment that new generation of scientists will demand.
   (cooperative framework such as ASTERICS are just the first step)

• They are pushing towards "interoperability" of Data and "heterogeneity" of
  Computing Architectures:
     -> a new paradigm integrating HPC + HTC + Data Archive

• They are requiring new profiles: Data Scientists.